



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

**“Media Bias in East Asia:
Framing of the Umbrella Movement
in English Language Online News Outlets”**

verfasst von / submitted by

Paul Balluff, B.A.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 864

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Wirtschaft und Gesellschaft Ostasiens UG2002

Betreut von / Supervisor:

Univ.-Prof. Mag. Dr. Rüdiger Frank

Abstract In the context of mediated public diplomacy and mediated discourses, this thesis investigates the similarities and differences in the reporting of the 2014 Hong Kong Umbrella Movement in East Asian English language newspapers. The basis is a corpus of $N = 2,197$ documents from nine English online news outlets that is analyzed with automated content analysis techniques. The analysis investigates the salience of the protests across countries, framing, and quotation patterns. The investigated framing devices are derived from the protest paradigm and relate to the news perspective, an overly emphasis on violence, propagating the protesters cause, and foreign influence on the protesters. Additionally, results by the Asian Barometer Survey are compared against the news media framing to gauge the diversity in opinions on the matter. The findings suggest that the newspapers framing does not exhibit strong differences within a single country. Exception are the South Korean newspapers that have opposing perspective on the Hong Kong protests. Furthermore, geographical and political proximity seem to have an influence on media salience of the protests. While newspapers from Mainland China exhibit a skeptical perspective on the protests, Taiwanese newspapers are generally in favor of the Umbrella Movement.

Zusammenfassung: Im Kontext von medialer öffentlicher Diplomatie und medialisierten Diskursen untersucht diese Masterarbeit Gemeinsamkeiten und Unterschiede der Medienberichterstattung in englischsprachigen, ostasiatischen Zeitungen über die Regenschirm-Bewegung im Jahre 2014. Grundlage bildet ein Corpus mit 2,197 Dokumenten aus neun englischsprachigen Onlinezeitungen, welcher mit automatisierten Inhaltsanalysetechniken untersucht wird. Die Analyse untersuchte die mediale Sichtbarkeit der Proteste in den Ländern, das Framing, und Muster in Zitierweisen. Die untersuchten Frames sind vom Protest Paradigma abgeleitet und konzentrieren sich daher auf die allgemeine Perspektive der Berichterstattung, die Hervorhebung von Gewalt, das Propagieren der Beweggründe, und den Einfluss externer Mächte auf die Protestteilnehmer*innen. Zusätzlich werden Ergebnisse des Asian Barometer Surveys zur Analyse hinzugezogen, um die Frames in der Medienberichterstattung mit dem öffentlichen Meinungsspektrum in Ostasien zu vergleichen. Die Ergebnisse deuten darauf hin, dass es keine starken Unterschiede in den Frames innerhalb eines Landes gibt. Die Ausnahme stellen die südkoreanischen Zeitungen dar, welche eine gegensätzliche Perspektive auf die Proteste in Hong Kong bieten. Weiterhin scheinen geographische und politische Nähe einen Einfluss auf die Sichtbarkeit der Proteste in den Medien zu haben. Die Zeitungen vom chinesischen Festland präsentieren

eine skeptische Perspektive auf die Proteste, wohingegen die taiwanischen Zeitungen den Protesten allgemein freundlich gestimmt sind.

獻給陳凱欣

CONTENTS

List of Figures	vi
List of Tables	vi
Nomenclature	vii
1 Introduction	1
1.1 Challenging Beijing	1
1.2 Research Question	2
1.3 Analytical Approach	3
1.4 Rectifying Names	4
1.5 Thesis Outline	5
2 Literature Review	6
2.1 Content Analysis	6
2.2 Concepts in News Media Research	11
2.3 Protests in News Media Research	14
2.4 The Umbrella Movement	19
2.5 Summary of Discussed Research	23
3 Analytical Framework	25
3.1 Sample Definition	25
3.2 Criteria	26
3.3 Data	35
3.4 Evaluation of Content Analysis Results	38
4 Empirical Part	41
4.1 Collected Text Data	41
4.2 Salience	42
4.3 Framing	44
4.4 Quotations	52
4.5 In-house vs Bought Copy	53
4.6 Public Attitudes towards Protests	56
5 Analysis	58
5.1 Salience	58
5.2 Framing	60
5.3 Quotations	63

5.4	In-house vs Bought Copy	65
5.5	Public Attitudes	66
5.6	Summary of Analysis	67
6	Discussion	69
6.1	Limitations	69
6.2	Country-wise discussions of newspapers	70
6.3	Protest Paradigm	73
6.4	Differences and Similarities between Countries	75
7	Conclusions	80
	Bibliography	82
	Manual Coding Rules	95
	Tables	96

LIST OF FIGURES

4.1	Search hits results as proportion of monthly total documents. Grouped by country and outlet.	42
4.2	Latent semantic scaling results as longitudinal plot.	47
4.3	Topic model optimization metrics.	48

LIST OF TABLES

3.1	Summary of criteria for comparison	27
4.1	Overview of collected documents and search string hits	41
4.2	Average document length by Newspaper	43
4.3	Average tonality of documents mentioning the protests	44
4.4	Top 20 positive and negative terms of dictionary generated with LSS	45
4.5	Summary of results for fitted LSS Textmodel	46
4.6	Overview of Keyword Assisted Topic Model Results	50
4.7	Summary of topic model based frame classification	51
4.8	Summary of quotation counts by actor groups	54
4.9	Summary of documents sourced from agencies	55
4.10	ABS: Overview of survey respondents choosing freedom to participate in protests as essential	57
4.11	ABS: Overview of Freedom and Liberty Score	57
5.1	Summary of analysis.	68
1	LSS Post-hoc test results territory comparison (Games-Howell)	96
2	LSS Post-hoc test results newspaper comparison (Games-Howell)	97
3	ABS Wave 3 Freedom & Liberty Index Post-hoc test results (Games-Howell)	98
4	ABS Wave 4 Freedom & Liberty Index Post-hoc test results (Games-Howell)	98
5	Summary of quotation counts by actor groups (full table)	99

NOMENCLATURE

Abbreviations

ABS	Asian Barometer Survey.
AFP	Agence France-Presse.
AP	Associated Press.
API	Application Programming Interface.
BBC	British Broadcasting Corporation.
CCP	Chinese Communist Party.
CUHK	Chinese University of Hong Kong. Chinese: 香港中文大學.
DPP	Democratic Progressive Party, Taiwan. Chinese: 民主進步黨.
HK	Hong Kong.
HKFS	Hong Kong Federation of Students. Chinese: 學聯.
HKU	The University of Hong Kong. Chinese: 香港大學.
HTML	Hypertext Markup Language.
KMT	Kuomintang, Taiwan. Chinese: 國民黨.
LDA	Latent Dirichlet Allocation.
LSS	Latent Semantic Scaling.
NYT	The New York Times
OCLP	Occupy Central with Love and Peace. Chinese: 讓愛與和平佔領中環 or 和平佔中.
POS tagging	Part-of-speech tagging.
PRC	People's Republic of China; China.
ROC	Republic of China; Taiwan.
RTHK	Radio Television Hong Kong, public television and radio station. Chinese: 香港電台.

SCMP South China Morning Post.

TVB Television Broadcasts Limited. A private television company in Hong Kong.

Names in Korean

Candlelight Demonstrations 촛불항쟁. Protests in South Korea, 2016–2017.

Chosun Ilbo 조선일보. Newspaper based in South Korea.

Hankyoreh Daily 한겨레. Newspaper based in South Korea.

JoongAng Daily 중앙일보. Newspaper based in South Korea.

Korea Times 코리아타임스. Newspaper based in South Korea.

Kyunghyang Shinmun 경향신문. Newspaper based in South Korea.

Names in Simplified Chinese

China Daily 中国日报. Newspaper based in China.

Global Times 环球时报. Newspaper based in China.

People's Daily 人民日报. Newspaper based in China.

Xi Jinping 习近平. President of People's Republic of China 2013–today.

Names in Traditional Chinese

Apple Daily 蘋果日報. Newspaper based in Hong Kong.

Benny Tai 戴耀廷. A former professor for law at the University of Hong Kong. One of the founders of the Occupy Central Campaign.

Chan Kin-man 陳健民. Professor for sociology at the Chinese University of Hong Kong. One of the founders of the Occupy Central Campaign.

Chu Yiu-ming 朱耀明. Reverend of the Chai Wan Baptist Church in Hong Kong. One of the founders of the Occupy Central Campaign.

Ken Tsang 曾健超. District council member who was severely injured by the Hong Kong police on October 15, 2014.

Leung Chun-ying 梁振英. Chief Executive of Hong Kong SAR (2012–2017).

Liberty Times 自由時報. Newspaper based in Taiwan.

Ma Ying-jeou 馬英九. President of ROC (KMT, 2008–2016).

Ming Pao 明報. Newspaper based in Hong Kong.

Occupy Central 佔中. Protest Movement prior Umbrella Movement.

Oriental Daily News 東方日報. Newspaper based in Hong Kong.

Scholarism 學民思潮. A student organisation in Hong Kong.

Ta Kung Pao 大公報. Newspaper based in Hong Kong.

Taipei Times 台北時報. Newspaper based in Taiwan.

The China Post 英文中國郵報. Newspaper based in Taiwan.

The Standard 英文虎報. Newspaper based in Hong Kong.

Tsai Ing-wen 蔡英文. President of ROC (DPP, 2016–today).

Umbrella Movement 雨傘運動. Also referred to as Umbrella Revolution (雨傘革命). Other common terms are derived from the Cantonese word for umbrella: 雨遮革命, and 遮打革命.

Wen Wei Po 文匯報. Newspaper based in Hong Kong.

1 INTRODUCTION

1.1 CHALLENGING BEIJING

“We’ll be back” was written in black letters on a large yellow banner that covered a street in Hong Kong Admiralty district. The banner was left at the protest site on December 10, 2014 before the last participants were evicted by the police (Lee and Chan 2018, p. 179; VOA News, December 10, 2014). Less than five years later, on June 9, 2019 over half a million people protested in Hong Kong against the government’s plan to introduce a new extradition law (Reuters, June 9, 2019). During the course of the year, the protests continued. Many of the participants wore yellow raincoats and carried umbrellas. These were the symbols of the civil disobedience campaign that choked important streets in the city in 2014 and gained international attention as the *Umbrella Movement*. The 2014 campaign was smaller with up to 100,000 participants but had similar values and characteristics (Ho 2019, p. 2). The majority were students who demanded universal suffrage for the 2017 Chief Executive elections without interference from Mainland of China. However, the demands of the 2014 protests were not, but the spirit of the movement continued. Since then, the issue increasingly received international attention and raised awareness for the implications of China’s advances.

In international news, the majority of protests go unnoticed, only few reach a critical mass and gain attention in foreign media. In certain cases, protest events in foreign countries arguably have an influence on domestic politics. In 2019 and early 2020, we could observe how the protests in Hong Kong added a new agenda item to the presidential elections in Taiwan. The public debate shifted from economic prospects towards security and autonomy. China’s tighter control on Hong Kong caused Taiwanese voters to reconsider cross-strait relations (Brown and Churchman 2019, p. 67; Copper 2020, p. 47; Templeman 2020, p. 86). Although this example is rather specific, it illustrates that foreign events do influence the public mood and can be utilized by governments to antagonize foreign countries. In the international mediated public arena, governments may aim at influencing foreign opinions on current events and steer the framing towards a narrative that presents their view on the story (Entman 2008, p. 90). These narratives can be used to justify outward actions such as sanctions or other interventions in foreign affairs, as well as the choice to openly support a foreign administration. Looking inwards, narratives can delegitimize oppositions, or justify policy decisions.

The rise of China raised strong international attention starting from the late

1990s (Zhang 2010, p. 254). And the Chinese government has intensified their efforts to improve the nation's image internationally through mediated public diplomacy (Chang and Lin 2014, p. 457). While Southeast Asians have a more positive attitude towards China due to the economic potential, the near neighbors such as South Korea, Taiwan or Japan are less positive towards China's growing power (Chu, Kang, and Huang 2014, p. 414). The East Asians have not reservations about China's economic impact on the region, but also in terms of security and territorial issues (p. 408). However, the views on China are diverse and polarizing at times, especially in terms of personal democratic values (p. 413).

The Umbrella Movement is embedded in a context of East Asian democratization (Ho 2019, p. 6). But to the same extend, it is also part of a resistance against China's rise (p. 17). Against this backdrop, the media coverage of the Umbrella Movement is subject to an international discourse that is influenced by outwards aiming public diplomacy and inward-looking political competition.

1.2 RESEARCH QUESTION

This thesis is concerned with patterns of reporting on the 2014 protests in Hong Kong in East Asian news outlets. East Asia has a diverse news media landscape with varying degrees of government control on reporting practices. Consequently, the narratives that they produce may show substantial differences, especially news outlets that are targeting foreign audiences. This thesis it aims to investigate common characteristics and differences in the framing of social protests across English news outlets from East Asia. Thus, the main research question is:

RQ: *How does the news coverage of the 2014 protests in Hong Kong in English news outlets from East Asian compare against each other?*

The differences and similarities in the news coverage about the 2014 protests are investigated. This thesis aims at identifying patterns of reporting that may be specific to only one news outlet or be shared by several outlets. The reporting patterns are specifically investigating a potential bias in favor or against the protesters. It also looks at framing devices that were used by the news outlets and to which extend they varied between them. The comparison considers differences between countries and within them by selecting two newspapers per country in order to investigate patterns of pluralism in mediated discourses. And finally, the thesis uses results from the *Asian Barometer Survey* to identify whether the perspectives presented in the newspapers match the public attitudes or differ greatly.

East Asian societies have diverse views on protest movements, especially when they are considered illegal. While the general populations in China, Japan, Hong Kong and Taiwan are considered to "frown" on protests, South Koreans tend to be more supportive of such activities (J.-E. Lee 2012, p. 412; Ho 2019, p. 8). Therefore, studies that investigate the news media framing of protests frequently contextualize the image presented in the media with

opinion surveys (Lee and Chan 2018, p. 26; Ho 2019, p. 87).

The relevance of this thesis lies in its attempt to examine the mediated discourse targeted at foreign audiences and to compare it against public opinion. Governments are intrigued to use English language news sources as a tool to disseminate their narratives for mediated public diplomacy (Cheng, Golan, and Kiouisis 2015, p. 744). These types of newspapers are also embedded in a business context, where economic agents seek for information, and thus rely on international news media for investment decisions (Tuchman 2002, p. 79). While the East Asian region showcases economic cooperation and peaceful diplomatic exchanges, there are frequently episodes where tensions become visible (Lai 2017, p. 179). This dualistic order of cooperation and competition (Foot and Goh 2018, p. 408) is expected to appear in the mediated arena as well. Given this context of mediated public diplomacy, this thesis can hint at strategic communication campaigns aimed at foreign audiences (Cheng, Golan, and Kiouisis 2015, p. 745). Also, it examines how much of the pluralistic discourses transcends national boundaries.

1.3 ANALYTICAL APPROACH

In order to compare the news framing of the Umbrella Movement across East Asian newspapers, first criteria for comparison are derived based on previous studies concerned with the framing of social protests in news media. Then a large corpus containing English news articles from Hong Kong, Taiwan, China, and South Korea is empirically investigated by employing an automated content analysis. For each country, the two largest¹ English language online news outlets are selected. Additionally, the findings of the analysis are also compared against results from the Asian Barometer Survey which feature question items relating to attitudes towards democracy (Lu 2012, p. 4).

Content analysis is a method that can disclose international differences in news content, expose propaganda techniques, and reveal the focus of attention (Krippendorff 2004, p. 45). As a research method, “content analysis helps to identify and describe patterns in media on a large scale” (Gunter 2002, p. 220). In the context of protest coverage, news outlets can choose to be biased in favor or against the protest participants. These choices are often deliberate and can point towards patterns where news outlets want to propagate the story from a certain perspective (Boyle and McLeod 2018, p. 301).

The automated techniques used in this thesis are text models. They are based on computational methods that turn the complexities of a written text into numerical data and use statistical estimations to make classifications and predictions. The results of the models are not perfect and as Box (1979, p. 202) said: “all models are wrong, but some are useful”. This thesis works with the assumption that the models are at least internally coherent, but the generated numbers are not exact. The directionality and comparison of the results is of greater importance. To add this aspect to the text models, *The Guardian* as an additional source that is not from East Asia is included in the corpus. Its in-

¹Ranked according to their size of readership.

clusion helps to create a reference point for the model results. This of course, does not imply that news stories of *The Guardian* are in any means a standard of reporting.

1.4 RECTIFYING NAMES

This thesis investigates the news media framing of the 2014 pro-democracy protests in Hong Kong. Using the definition by Ho (2019, p. 62), “a protest is operationally defined as a public and confrontational act mounted by an intentional group of people whose claim involves a conflict of interest with their opponents.” There were several protests and groups active during that year advocating for universal suffrage for the 2017 Chief Executive elections. The protest activities began in 2014 as early as in February when the student organization *Scholarism* stated their demands for universal suffrage (p. 121). Throughout March until September, protest activities related to these demands were conducted by various groups, most notably by *Occupy Central with Love and Peace* (OCLP) (p. 122). Only when the demonstrations escalated on September 28 and the protesters used umbrellas to shield themselves from pepper spray, the news media labelled the protests “Umbrella Movement” (Lee and Chan 2018, p. 77). Hong Kong’s Umbrella Movement usually refers to a diffuse group of protesters who occupied parts of the city from late September until mid-December 2014 to advocate for universal suffrage (p. 1). In the context of this thesis, the term “Umbrella Movement” is used to describe all protest activities and groups of 2014 that were advocating for universal suffrage. The labels “Occupy Central”, “protests in Hong Kong”, “2014-protests”, and “Umbrella Movement” are all here used interchangeably.

Also, the scope of this thesis is restricted to the year 2014. By default, references to months or days without explicitly mentioning the year, always refer to 2014.

Various East Asian names from different languages appear in this thesis. While there are numerous methods for romanization, this thesis uses the conventional spellings of names as they appear in the English media. For example, the names of politicians from Hong Kong, mainland China and Taiwan, are used as they are commonly found: Xi Jinping (instead of *Hsi Chin-ping*), Tsai Ing-wen (instead of *Cai Yingwen*), or Leung Chun-ying (instead of *Lang Zhenying*). In case there is no canonical romanization available, the dominant romanization system of the country where the name originates from is used. To clear up any potential confusion, there is a glossary of non-English terms available with the romanization used in the thesis.

Finally, this thesis often refers to political constructs such as the Hong Kong SAR as a “country”. While it is technically not correct, this label is used for convenience.

1.5 THESIS OUTLINE

The literature review provides an introduction on what a content analysis is, its assumptions and common methods. Then it reviews specific techniques that are used in this thesis. Next, an overview of studies that are concerned with the coverage of social protests in news media is presented. The results of the literature review are used to create an analytical framework that contains criteria for the comparison of different newspapers. The framework is then applied, and the results are presented. After the comparative analysis of criterion, the subject matter is discussed. Finally, a conclusion will summarize the findings and point at potential follow-up research.

2 LITERATURE REVIEW

First, an introduction to the applied methodology is presented. The section reviews canonical works about content analysis and their recommended procedures for analyzing bodies of text. It also touches on basic assumptions of content analysis as well as presents approaches for automated content analysis.

Next, two concepts from journalism and communication studies are explained that are relevant for this thesis. The first one being news media framing and the second is the protest paradigm. Then, a review of studies that are concerned with protest research based on content analysis techniques. Finally, this review is further focused on recent works about the Umbrella Movement in context of the news media.

2.1 CONTENT ANALYSIS

2.1.1 *Definition*

Content analysis is generally a “quantitative assessment of media output” that is applied to enquire a variety of research interests (Gunter 2002, p. 219-220). This research tool can be applied on any type of media, such as political speeches, films, novels, academic journals, paintings, social media, and so forth. In a context related to news media, researchers are often interested in investigating agenda setting effects, as well as the examination of patterns in the coverage of social groups, events or issues (p. 220). Krippendorff (2004, p. 18) defines content analysis as “a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use”. It is important to note here that this definition includes that the results are also reliable. Reliability is strongly connected to replicability, which entails that the research is systematic and subject to rules that ensure the findings are “divorceable” from the respective researchers (p. 18). The research process also has to withstand external criticism and has to be tested for validity (p. 19). Therefore, the interpretive schemes have to be explicated to reduce the potential number of findings that can be inferred from the data using the same research method (p. 24). That includes stating the particular context of the data and formulating questions that are directed at the data (Jensen 2002b, p. 8).

Quantitative analysis of newspaper texts is a subcategory of content analysis, and therefore it follows the same basic principles as its parent technique. Specifically, the domain of news media is sensitive towards implicit norms and expectations towards the news media and journalism. The normative aspect of analyzing news media appears in at least two different ends. Firstly, the ana-

lyzed data is embedded in a normative context, the authors of news have values and norms that are embedded in it. Secondly, the researchers can design their questions with certain normative expectations in their mind (Krippendorff 2004, p. 28). From an analytical perspective, using a normative approach does not make the findings less valid, however these assumptions have to be explicated to ensure the findings can be validated. However, applying normative concepts also poses the risk of producing research that is embedded in a historical and cultural context that can thus be difficult to compare to further studies. Inferring content from text thus requires the research process to be as transparent as possible by explicitly describing all concepts that are relevant for the study (Wimmer and Dominick 2011, p. 50).

2.1.2 *Constructivist viewpoint*

Krippendorff (2004, p. 25) generally follows a constructivist view for content analysis and argues that true objectivity can never be achieved. Murdock (2002, p. 42) also stresses that media as a form of communication constructs a “social reality”. Tuchman (2002, p. 87-88) follows suit in arguing that a constructivist mindset is the mode of choice for content analysis of news media, but adds to the discussion that the journalist and news organizations as agents are also embedded in a context of socio-economic circumstances that are reflected in the texts. Hence understanding findings of a content analysis, also relies on examining “political-economic preconditions, its organizational enactment, and its textual articulation” (p. 88).

Given the assumption that the research process itself uses social constructs, a sensible method to enhance the validity of the findings is by establishing a clear frame of reference which is used to compare the results of the content analysis to (Schröder 2002, p. 101). The sources of comparisons can include “real world” statistics, other forms of media, or content from a different time period. Some examples would be to compare the representation of female politicians in the news media against their proportion in the parliament, comparing the coverage of newspapers with television news, or comparing the change of the same source over a longer period of time (p. 101-102).

These recommendations suggest that constructivist thinking is useful for designing the research process, because this makes it easier to replicate and criticize the research against different cultural backgrounds or points in time. A research process that made sensible assumptions in a particular context in the past, can best increase its compatibility with future research if all steps of the process are documented in detail (Jensen 2002c, p. 256). However, this constructivist attitude towards the research process does not necessarily require that the research question or other parts of the planned study has to follow such a philosophy.

2.1.3 *Types of inferences*

Wimmer and Dominick (2011, p. 119) state that qualitative and quantitative content analysis are inherently different in nature. They make their distinction at

the level of logical inferences, where they argue that quantitative content analysis is deductive and qualitative is inductive. This suggestion is rooted in the level of flexibility used in the research process. The qualitative process uses open questions that can change throughout the process and guide the researcher to approach the content. Therefore, this type of research process starts at the particular and leads to conclusions of the general. The quantitative content, however, uses a standardized and static set of questions that are applied equally to each unit of analysis; hence it applies a general rule to particular observations (Wimmer and Dominick 2011, p. 118).

Krippendorff (2004, p. 42) suggests that there is no such distinction necessary, because any quantitative content analysis is an aggregation of a large amount of qualitative inferences. He proposes to categorize flexible approaches as other types of textual inferences such as discourse analysis or rhetorical analysis (p. 16). Furthermore, the difference between inductive and deductive is merely a theoretical distinction, because in research practice most content analyses follow an *abductive* inference (Jensen 2002c, p. 264). Krippendorff (2004, p. 85) highlights that abductive inferences require a *warrant*. Researchers often use “a mixture of statistical knowledge, theory, experience, and intuition” (p. 38). Meaning that research is built upon an array of external knowledge that is applied in its inductive or deductive reasoning. This can include knowledge about the content, language, history, or previous research on that matter.

Baden and Stalpouskaya (2015, p. 5) note that practical large-scale content analysis, follows a combined strategy that “enable[s] a mapping of inductively collated patterns [...] onto a range of deductively defined criteria”. This would present a practical research process that can leverage the advantages of flexible qualitative inquiry to formulated categories for later quantitative examination (p. 5). While they do not explicitly dwell too deep in this subject and do not use the term “abductive”, the argument follows the recommendations by Krippendorff (2004, p.38) and D’Angelo (2018b, p. xxxii).

The inferences for this thesis are derived from previous studies by looking for the existence of certain frames in the corpus. This renders the nature of this study as deductive.

2.1.4 Steps of newspaper analysis

A quantitative analysis of newspaper content can be broken down into six distinct steps. The first step is usually to define the unit of analysis, meaning the elements that are being studied and quantified. For newspaper analysis that would usually be the textual data on the article level. However, other segmentations are also plausible, such as the sentence or paragraph level. Furthermore, additional information such as the images accompanying articles can be taken into account as well (Gunter 2002, p. 220).

The next step is to define the sample, where Krippendorff (2004, p. 83-84) suggests several parameters: time frame, sources, quantity, and composition. The time frame of the sample needs to have clear boundaries. The source selection includes the newspapers as well as the origin of the sample. This ranges from manual clippings of print sources to digital archives. Typically, a reduc-

tion of the total available data is necessary. This can be achieved by creating a subset based upon categories defined by the researcher or the medium. It is common practice to sample the data by keyword search matches (Gunter 2002, p. 222). For example, if the study is about the coverage of solar power technology, the researchers might choose to select material that contains matching keywords. Or it would also be reasonable to select based on newspaper sections and only include articles that correspond to a particular section, such as *economy*. In either case, it has to be an informed decision and the reduction of the sample is a compromise between staying close to the research question and making the research practically feasible.

The third step is to ensure a consistent structure of the data, so that analytical procedures can be applied to it (Krippendorff 2004, p. 84). For paper-based newspaper articles, it is often necessary to digitize it and to create a coherent structure. Conceptually, this applies for already digitally available data sources as well, because often the meta information across sources has to be standardized. Additionally, this step entails an array of sub-steps that are highly relevant for automated techniques. However, these pre-processing steps vary greatly depending on used techniques (Welbers, Atteveldt, and Benoit 2017, p. 4). The detailed steps that are relevant for this study are explained below. But to give a brief overview, there are some common type of pre-processing steps that are shared amongst most techniques such as removal of redundant white spaces, URLs, or special characters.

The fourth step is where specific content analysis techniques are applied to the texts and as a result the information of the data is reduced (Krippendorff 2004, p. 84-85). The intricate details of the applied techniques can greatly vary; however, they should be grounded in concepts that relate to the research question. Generally, an abstract concept of the research question is broken down into single variables that measure the presence of said concept. Then, the variables are *coded* from the text according to a set of rules. A common method would be to count the occurrences of selected terms (or tokens) and compare the frequency against other data points. For example, by quantifying the search keyword "terror*" in various newspapers and map that across time one can identify the changing prevalence of the issue throughout the past decades (Welbers, Atteveldt, and Benoit 2017, p. 11). Of course, there are also more complex methods possible, such as coding texts for the presence of emotional language, or tonality (Grimmer and Stewart 2013, p. 275). The desired result is an aggregation of the data with a loss of detail, but efficient representation that is relevant for the research.

The fifth step is to "move an analysis outside the data" (Krippendorff 2004, p. 85). Meaning that by making abductive inferences, the aggregated data is contextualized. The content analyst uses any additional information available to interpret the findings. For example, the frequency of the term "terror" occurring in a newspaper can be compared to actual acts of terror in the same time period and geographic region. This would enable to enquire if the term appears proportional to events (Baden and Stalpouskaya 2015, p. 14).

The final step is to narrate the findings by trying to answer the proposed research question (Krippendorff 2004, p. 85). Ideally, it would also highlight,

how the content analysis added value to the field of research and what future studies could be conducted (Krippendorff 2004, p. 85).

The steps outlined above are in a logical progression. However, it is possible to proceed in a looped fashion at any point. Therefore, it is legitimate during practical research to return to a previous step in order to refine the results of the following (Baden and Stalpouskaya 2015, p. 14; Krippendorff 2004, p. 86).

2.1.5 Automated Approaches

As hinted above, the coding methods for variables can be distinguished into two types: manual or automated. The former one usually refers to coding methods where humans read document by document and manually fill in a set of variables. Automated approaches on the other hand use computers for this step and humans do not read every single document in the corpus (Trilling and Jonkman 2018, p. 159). The decision which values are set for each variable in every document, does not necessarily remain with the computer and is often strongly guided by the researcher. A brief overview of three commonly used automated techniques and their working principles are presented here.

The probably oldest automated technique relies on Boolean search strings and keyword dictionaries (Boumans and Trilling 2015, p. 11). For example, to measure how often Xi Jinping is mentioned in various news outlets, a simple search string could consist of the actor's name, such as: "Xi Jinping". The search criteria can be enhanced by including the actor's political function ("Chinese President") or adding different spellings ("Xi Jin-ping"). This yields a combination of search keywords that all relate to the same entity and are connected into a long search term with logical operators:

"Xi Jinping" OR "Xi Jin-ping" OR "President Xi" OR (President AND China)

This principle can be expanded to complex collections (*dictionaries*) of search keywords that also measure abstract concepts such as sentiment, or public issue areas such as climate change (Welbers, Atteveldt, and Benoit 2017, p. 255).

So called "bag-of-words" approaches make statistical analyzes based on the frequencies of all terms across all documents in the corpus (Boumans and Trilling 2015, p. 11). These techniques require the corpus to undergo a series of transformations. First each document is split into *tokens*, a smaller unit of text such as words or word stems (Welbers, Atteveldt, and Benoit 2017, p. 246). Then the frequency of each token in each document is counted in a document-term-matrix (p. 252). This numerical representation allows supervised and unsupervised algorithms to compare the similarity of documents based on the occurrence of terms. Using meta information such as news outlet or date, distinct features of a group of documents can be compared to the entire corpus. This matrix representation of a corpus also enables to find co-occurrence patterns of terms (Boumans and Trilling 2015, p. 15). Topic modelling is a common type of unsupervised techniques that use co-occurrence and is utilized for automated frame coding. The special characteristics of these approaches is that they reduce the corpus and let it

speak for itself without employing pre-defined categories for analysis (Blei, Ng, and Jordan 2003; Blei 2012; Boumans and Trilling 2015, p. 15). One disadvantage of bag-of-word techniques is that they discard the context of the terms and thus there is a loss of semantic information (Welbers, Atteveldt, and Benoit 2017, p. 252).

Word embedding techniques aim at preserving semantic information associated with words. These types of techniques describe each word in relation to the previous and following ones, and thus enables to train a language model on a large corpus. Mikolov et al. (2013) developed a method that is capable of representing words in a vector space with 640 dimensions. The model is capable to represent word relationships such as “man” relates to “king” as “women” relates to “queen” (p. 2). Once the language model is trained, it can be used in a variety of tasks, and most importantly to encode words into a numerical input for learning-based algorithms.

This summary is by far not exhaustive, new frontiers in machine learning enriched the available techniques substantially (Boumans and Trilling 2015, p. 12). Supervised models rely on training a classifier based on a large sample data set of human annotated data (p. 13). Further, semi-supervised techniques take a small sample of user input and scale it automatically to fit the available data set (Watanabe 2020, p. 4).

2.2 CONCEPTS IN NEWS MEDIA RESEARCH

2.2.1 News Media Framing

Content driven research of news media frequently investigates the frames used in journalistic texts (Carragee and Roefs 2004, p. 214). Framing is a communication process that “involves *selection* and *salience*” (Entman 1993, p. 52). Meaning that certain aspects of a “perceived reality” are made more visible to promote a particular presentation of it (p. 52). Thus, frames can be defined as “patterns of selection, emphasis, interpretation, and exclusion” (Carragee and Roefs 2004, p. 216). Entman (1993, p. 52) defines four functions of frames: problem definition, causal interpretation, moral evaluation and proposed treatment or solution. Frames become evident in text by the “presence or absence of certain keywords (or) stock phrases” (p. 52). He emphasizes that a not every sentence contains frames, and that a frame does not necessarily perform all functions at once. A typical contemporary example would be the terms used to refer to SARS-CoV-2. US President Donald Trump frequently calls it “China virus”, which suggests a problem definition (i.e. the virus originates from China).

D’Angelo (2018a, p. xxiv) points at the complex interplay of journalists, sources, and audiences. Where each party affects the conditions under which a message is constructed and received. Typically, journalists would rarely convey the message directly as it was framed by the source. Often, media workers implement an interpretative process that changes the framing depending on a complex array of factors. Additionally, the audience has their own specific associative patterns for words, that make the reception of a particular frame heterogeneous within the audience (p. xxix). Referring to the example above,

the frame “China virus” would also plausibly provide a causal interpretation for certain audiences (i.e. it is China’s fault that there is a virus). The media workers therefore might deliberately decide to use the provided frame as it is presented or chose a different presentation. However, frames are not always manufactured or reproduced to consciously influence public opinion. In some cultural or issue specific contexts, the audience also expects a certain degree of framing from the journalists, because framing also reduces complexity and assists the reader to comprehend the issue at hand faster. However, there are often situations where various sources establish different frames for the same piece of information. This interaction and competition of frames can make the perceived reality malleable for the message recipients (D’Angelo 2018a, p. xxv).

Frames are contested due to their potential to influence public opinion (Wanta, Golan, and Lee 2004, p. 374). Frame contestation is a continuous process amongst several actors who aim to make their interpretation of reality the dominant one (Carragee and Roefs 2004, p. 219). The frames that are present in the news media are the “true imprints of power”, because they showcase which actors succeeded in propagating their frame (Entman 1993, p. 55). The contestation ranges on a continuous scale from total dominance of a single frame to a frame parity where multiple frames are more or less equally present (Entman 2008, p. 90). The contestation of frames can also vary in their scope. While it is most commonly understood that the struggle for frames in the news media happens within a political entity, it is also a struggle that reaches beyond national boundaries. States use mediated public diplomacy to propose their set of frames to the public discourse in foreign nations (p. 90).

News stories often cover topics with a certain story angle, and researchers pay attention to these aspects of the news media. Carragee and Roefs (2004, p. 217) argue that framing research has blurred the lines between typical characteristics of news media. They argue that a story angle alone would not qualify to “construct particular meanings” as frames would (p. 217). They emphasize the importance of the ideology and power relationships residing in the frame and argue for finding “meaningful definitions” (p. 218). This calls for a nuanced view to distinguish various types of frames (or *framing*). In more recent research, the blurriness of framing has been reduced by distinguishing between *generic frames* which are universally present in all types of news coverage, and *issue specific frames* that can vary greatly across issues (Semetko and Valkenburg 2000, p. 95; Burscher et al. 2014, p. 203). The generic frames that are commonly evaluated are “human-interest”, “conflict”, “economic consequences”, and “morality” (Burscher et al. 2014, p. 193). The human-interest frame is concerned with the depiction of individuals affected by the covered topic and brings an emotional angle to the piece (Semetko and Valkenburg 2000, p. 95-96). A conflict frame is present if the news item emphasizes the struggle between two or more parties on the subject at hand (p. 95). The economic consequence frame focuses on a potential financial impact for individuals or groups (often taxpayers). And the morality frame is concerned with evaluating ethical aspects of an issue, often via a proxy such as an interview or comment from a public figure (p. 96). An important characteristic of the generic frames is that they all have *inherent news value*, meaning that each of them is deemed

interesting for the audience (Burscher et al. 2014, p. 196).

Carragee and Roefs (2004, p. 218-219) further differentiate frames from attributes and issue positions. Attributes provide a direct evaluation that is often expressed through adjectives in text and hence constitute a certain tonality. Issue positions can be a representation of single or multiple viewpoints on a given subject. Attributes and issue positions both make news items opinionated, but their presence is independent from the presence of frames. Frames in contrast, describe the way *how* the story is presented and which “ideological packages” are employed to construct a particular meaning (Lee and Chan 2018, p. 152). Vreese and Boomgaarden (2003, p. 364) suggested to bring the concept of attributes and frames together in *valenced frames*, where they argue that frames have an inherent valence to them. For most frames this valence can be a polar quality and thus these frames suggest positive or negative aspects (p. 363). For example, the economic consequences frame can be presented positively, neutral, or negatively (p. 368).

One commonly observed result of news framing is a certain bias that news outlets present. News media bias can be understood as a preference of using a certain set of frames. Eberl, Boomgaarden, and Wagner (2015) provide an approach that enables operationalization of news media bias. They distinguish between three types of bias: visibility, tonality and agenda (p. 1127). Visibility bias describes how often an actor appears in a given media outlet (p. 1128). The tonality bias measures the evaluation of said actors and considers the wording surrounding them. Positive or negative context of coverage can influence the consumers evaluation of the actors and is the qualitative counterpart to visibility bias (p. 1128). The final type of bias considers which issues are discussed in the media and who is in control of choosing the issues. Depending on the configuration of the media system and the orientation of the news outlet, it can either be the editor’s “final say” or an external influence that decides on the issues discussed (p. 1128). Eberl, Boomgaarden, and Wagner (2015, p. 1129) measured this type of bias by analyzing the issues raised in party press releases and comparing that to the issues presented in the political coverage. Regarding the effects of media bias the study by Eberl, Boomgaarden, and Wagner (2015, p. 1126) agrees with previous works (such as Belle 2010, p. 92) that “salience” (or visibility) has a stronger influence than tonality.

2.2.2 Protest Paradigm

The framing of protests in the news media is frequently examined with content analysis techniques. Most studies investigate forms of systematic bias towards protest movements. One recurring pattern of negative media bias is the *protest paradigm* (Chan and Lee 1989, p. 98). It describes a tendency in the news to focus on specific aspects of the protest, mainly violence and disruptions, but neglecting the issues and arguments raised by the protesters (Lee and Chan 2018, p. 31). Journalists relying on the paradigm would show a preference towards quoting official government sources rather than protest participants (p. 31). According to the paradigm, this type of selectivity and framing intentionally and unintentionally undermines and delegitimizes protests. This pattern has been

observed in a variety of geographic and historical contexts, with the notion that the more a protest challenges the status quo in a society, the more likely the journalists will rely on the paradigm (McLeod and Detenber 1999, p. 4). Another important factor that influences the framing of protests is the political alignment of the news outlet. Unsurprisingly, more ideological congruence between a protest and a news outlet leads to a more favorable coverage (Kim and Shahin 2020, p. 402). Journalists may choose to strengthen the protesters by quoting them, highlighting their intentions, and providing background information on the issue at hand (McCluskey et al. 2009, p. 542).

The ideological overlap transcends borders, and research suggests that newspapers tend to be more sympathetic with the protesters in foreign countries, especially if the protest is ideological well aligned with the news outlet (Ha and Shin 2016, p. 551; Kilgo et al. 2018, p. 624; Kim and Shahin 2020, p. 404). In the sphere of international mediated public diplomacy, governments may attempt to influence public opinion on foreign protest movements (Entman 2008, p. 90; Hartig 2016, p. 656). Based on the current interests an administration, it may choose to issue statements in support of protests in foreign countries in order to undermine the legitimacy of the corresponding government. Or it might also be part of the public diplomacy strategy to support the foreign government and to present the protests in an unfavorable manner (Ha and Shin 2019, p. 213). It is part of many journalistic routines to include statements of the domestic government on foreign events, usually to make the foreign issue more relatable to the local readership (p. 213). While not all research employs the protest paradigm, it is still a useful concept, because it provides a more or less consistent set of frames that can be used for comparison of various studies.

2.3 PROTESTS IN NEWS MEDIA RESEARCH

2.3.1 Overview

Researchers have used content analysis to investigate the coverage of civil movements and protests in newspapers for a long time (e.g. Chan and Lee 1989). Shahin et al. (2016, p. 146-147) suggest classifying such research into three categories: idiographic, transnational, and comparative studies. The first one, considers the domestic coverage of local protests. For example, the idiographic study by Feng (2017) who compared the coverage of the Umbrella Movement in four local newspapers in Hong Kong. Transnational studies analyze the foreign coverage of a domestic protest, such as done by Ojala and Pantti (2017) who compared the framing of the Euromaidan Protests in four European Newspapers through a manual content analysis. And comparative studies include both foreign and domestic coverage of domestic protests and examine the differences across nations and news outlet orientation. These three categories can also gain another quality by considering whether the comparison is synchronous or diachronous. In the following section, I will provide an overview of various recent studies that employed a content analysis to examine the media coverage of protests in the 2010s. This overview is not exhaustive, but almost all works

have some relation to East Asia, by either investigating an event in East Asia, or by including a news outlet from East Asia. Most of these works are based on manual content analysis techniques, therefore, I included selected studies that used automated approaches with a large sample.

2.3.2 *Arab Spring*

Harlow and Johnson (2011) manually analyzed the content of 208 documents from traditional and non-traditional online media that covered the Egyptian protests in early 2011. Investigating the prevalence of the protest paradigm in social media, the sample compared the framing and style articles of the *New York Times*, posts of a journalist's Twitter account, and entries in *Global Voices* a citizen journalism blog (p. 1363). The study suggests that the social media and blogs more frequently favor the protesters and preferably quotes them. The mainstream outlet, however, relies more on government statements, and employs drama and spectacle frames. But the traditional newspaper also features richer coverage of background information (p. 1367).

Karyotakis et al. (2017) compared the framing and news sources in the on-line editions of *Al Jazeera*, *China Daily* and the BBC. The sample included 92 text documents that covered the days surrounding President Hosni Mubarak resignation (p. 69). The investigated frames include human-interest, conflict, economic consequences, public affairs, and civil unrest (Karyotakis et al. 2017, p. 71; compare: Burscher et al. 2014, p. 193). The public affairs frame has a blurry definition and "focuses mostly on political actors or authorities", meaning that it describes the journalists' reliance on official information (Karyotakis et al. 2017, p. 69-70). The civil unrest frame describes unusual behavior of citizens in the wake of social breakdown such as looting of shops (p. 70). *Al Jazeera* most prominently featured this civil unrest frame, and only rarely used the human-interest frame. The BBC relied strongly on official sources, but also employed the human-interest frame to a moderate extent. *China Daily* almost entirely relied on official sources and did not emphasize individual stories frequently. Interestingly, *China Daily* used less frequently the civil unrest frame than BBC or *Al Jazeera* and featured as only outlet an economic consequence frame in the sample (p. 71-72). Additionally, there are traces of a foreign influence frame and an anti-American stance present in *Al Jazeera* and *China Daily*. However, while *Al Jazeera* mainly used in-house journalists for their news articles, *China Daily* heavily relied on news wires provided by *Agence France-Presse* (AFP) (p. 72).

As part of a follow-up study, Ha and Shin (2016, p. 538) investigated the framing of the Arab Spring in South Korean newspapers and how the foreign issue is "domesticated" according to the newspapers' respective political orientation. They specifically focused on frames that present the causes of the protests, potential solutions, and contextualization of South Korea's national interests in the Arabic World (p. 542-543). Based on a small *N* and inductive pre-study (see: Ha 2015, p. 746), each dimension was broken down into six pre-defined frames (e.g. the dictatorship is the cause of the protests). They collected 1,012 news stories over a period of 18 months from the conservative

papers *Chosun Ilbo* and *JoongAng Daily* and the liberal newspapers *Hankyoreh Daily* and *Kyunghyang Shinmun* (Ha and Shin 2016, p. 543). They manually analyzed a random sample of 506 news stories according to the eighteen pre-defined frames in a dichotomous fashion (present / not present). Their results suggest that the political orientation of the newspaper significantly determines the presentation of potential solutions to the conflict, as well as the contextualization to national interests (p. 551). Especially, the liberal papers were critical of the U.S. involvement in the Middle East, while conservative papers highlighted the potential for uprisings in PRC and North Korea (p. 552).

Du (2015) compared the framing in newspapers from Hong Kong, China and Taiwan with a specific focus on the framing of internet censorship in the Arab Spring. They conducted a manual content analysis of 108 Chinese language news stories ranging from December 2010 until March 2012. The sample was collected via search keywords from an online news archive and included 19 newspapers from China, 13 from Hong Kong and 7 from Taiwan (p. 105). They coded generic frames mentioned above, and issue specific frames on news perspective, government action, protester behavior and internet censorship. The news perspective was coded on an interval scale ranging from anti-protester to anti-government. Analogous, the government action ranged from peaceful conciliation to violent suppression, and the protester behavior from negative to positive (p. 106-107). Framing of internet censorship was assessed through dichotomous variables in terms of mentioning one of the following items respectively: obstacles to internet access, limits on available content, violation of internet user rights, influence of the internet on Arab Spring, and an anti-censorship perspective (p. 110). The results suggest that the Hong Kong and Taiwan based newspapers tend to have a favorable perspective on the protesters, evaluated their behavior as generally peaceful, and presented government action as suppressive (p. 108-109). The Chinese media tended to frame the events in the opposite direction, and frequently quoted government officials as well as attributed the instability to economic grievances (p. 110). Hong Kong newspapers were most likely to report on internet censorship and shutdown of mobile networks, while Chinese newspapers framed it as a general breakdown of communication or rarely mentioned the issue at all (p. 109-111).

In the context of China's efforts to become a significant player in the international news media market, Ha and Shin (2019, p. 203) analyzed the framing of the Arab Spring in *China Daily*. The study compares front page news stories published in *China Daily* written by Western news agencies with original content produced by in-house journalists. They manually analyzed 273 news items from December 2010 until October 2011 for four criteria (p. 207-209). The variables included the origin of the news story (in-house, or agency), nine different news frames, the object of responsibility (e.g. the ruling governments, the West, the opposition), and the referenced news sources (e.g. ruling governments, Chinese government). The nine deductively constructed frames include a Western imperialism frame, a social stability frame, a Chinese national interest frame, a China's active role frame, a political freedom frame, and a suppression frame (p. 206). The results of the content analysis suggest that the news items written by Chinese journalists are significantly more negative about the

Arab Spring than those news items from international agencies. The Chinese authored articles generally delegitimize the protests by emphasizing the importance of social stability, relying more on Chinese government statements, and marginalizing the political demands of the protesters (Ha and Shin 2019, p. 210). They also tend to blame “the West” more often than the ruling governments for the insurgencies, while highlighting the Chinese governments peaceful means of non-intervention (p. 212).

2.3.3 *Sunflower Movement*

Chiao and Huang (2017) conducted a small-scale study (N = 27 articles) where they compared the coverage of the *Sunflower Movement* in two Taiwanese newspapers (*United Daily News*, *Focus Taiwan*) and the *New York Times*. They specifically investigated the lexical preferences, syntactic choices (e.g. proportional usage of active voice), and the usage of quotations (p. 32). These three open-ended indicators gauge the news outlets’ framing of the events. While the Taiwanese newspapers focus on conflict and support of the status quo, the foreign coverage has a more neutral stance by stating the events without adjectives and consistently explaining the reasons for the protests (p. 40). Thus, the results of the study point towards the protest paradigm, especially with the nuance that local journalists rely on it more than foreign ones. While this in depth look on a small sample compares the three newspapers against each other, it does not provide insights how the coverage compares against the newspapers themselves. It remains unanswered for example, whether the *New York Times* generally prefers to use the active voice to compose their stories.

2.3.4 *Euromaidan Demonstrations Ukraine*

There are several studies on news media coverage of the developments in the Ukraine between late 2013 and mid-2014. However, none of them investigates the Euromaidan Demonstrations in an East Asian context. But it is still useful to review a small selection of works that employed automated techniques for content analysis, especially considering that the Ukraine protests lasted for a longer period of time and thus the studies presented have a longitudinal aspect as well.

Pashakhin (2016) used a topic modelling approach to compare the framing of Ukrainian crisis in Ukrainian and Russian TV channel websites. The sample contained over 50,000 written news articles starting from September 1, 2013 an October 1, 2014 and was analyzed using the topic modelling approach LDA (p. 104). The results suggest that the Russian news did not report much about the Euromaidan movement, while the Ukrainian news framed the protesters as terrorists (p. 104).

Watanabe (2017a) used an automated approach to analyze the framing of the *democracy* and *sovereignty* in the state-owned Russian news agency TASS. He investigated consistent patterns in the framing of news wires by the state-owned agency by comparing it to the commercial agency *Interfax* (p. 227). The sample contains 103,236 stories from *Interfax* and 87,725 from TASS and ranged

from January 1, 2013 until April 21, 2014 (Watanabe 2017a, p. 229). Watanabe (2017a, p. 226) argues that “objective reporting” is a hard to define concept, because there are various sources of bias. Especially, considering that most studies are founded on Western-European or North American media systems, news bias and negative framing are concepts that can vary greatly based on journalistic norms and traditions (p. 226). This leads to methodological issues, which he attempts to resolve by using the news stories released by *Interfax* as a benchmark unit. Given that both agencies are Russian organizations, they are expected to have a very similar reporting focus and journalistic values. This cancels out the specific configuration of the Russian news media system and thus allows to analyze the reporting bias of the state agency (p. 226). This approach renders news framing and bias as a relational concept. First, the sample was filtered by using a geographical classifier that identifies the location of the news story (Watanabe 2017c, p. 2; see: Watanabe 2017b). The data was further enriched by classifying the amount of text that is enclosed in quotation marks. In the next step, a dictionary for positive and negative framing of the keywords “democra*” and “sovereign*” was automatically created based on a reference corpus consisting of news wires from Reuters ($N = 21,718$) (Watanabe 2017a, p. 230). The dictionary contains sentiment scores for words related to democracy and sovereignty and was created through *Latent Semantic Scaling* (LSS), a word-embedding based semi-supervised machine learning technique (Watanabe 2017c, p. 3; 2020, p. 7). The performance of the dictionary was validated against a sample of 30 manually annotated news stories, where each sentence was coded on a five-point scale (Watanabe 2017c, p. 3). This procedure yielded a content analysis of the positive and negative framing of the Ukrainian sovereignty and democracy (Watanabe 2017a, p. 230). The news bias in TASS was estimated using a time series regression model that compared the differences between pivotal events (p. 230). The results suggest that TASS framed events more positively that are desirable to the Russian government. Especially, the positive framing of Ukraine’s sovereignty turned negative during the height of the protests which were directed against the pro-Russian government (p. 234). Another important finding is that the strongest framing effect was caused by quotations. Mentions of Russian entities together with quotes lead to a more negative framing of democracy in Ukraine (p. 234).

2.3.5 Candlelight Demonstrations

In a recent study, Kim and Shahin (2020) analyzed 387 articles of two Korean newspapers (*Chosun Ilbo*, *Kyunghyang Shinmun*) and two US dailies (*Wall Street Journal*, *New York Times*) in order to investigate transnational patterns of the protest paradigm. They specifically assessed the framing of the *Candlelight Demonstrations* in South Korea from November 2016 until April 2017 and compared the influence of the news outlets’ ideological affinities on the framing (p. 402). The manually coded framing devices were inductively derived by a small-scale pre-study and include cause, source, violence, size and gaze (p. 398). While the variables for cause, source, and violence correspond to previously mentioned studies, size and gaze are concepts that were not used previously.

“Size” is concerned with the presentation of the protest as large or small, where only explicit statements were considered (Kim and Shahin 2020, p. 398). “Gaze” describes how the news reports contextualize the protests in regard to the US-South Korean relations as well as the stability of the entire peninsula and region (p. 398). The results suggest that the liberal publications *Kyuhunghyang Shinmun* and the *New York Times* converged in presenting the protests as peaceful and large (p. 403). The Korean liberal paper quoted protesters most frequently, and the *New York Times* in a few instances as well, but still relied strongly on official statements (p. 401). The conservative papers *Chosun Ilbo* and the *Wall Street Journal* were similar in their presentation, however the US paper remained more balanced and avoided judging the scale of the protests and the behavior of the participants and also did not quote any of them (p. 403). The US newspapers converged in mentioning and explaining the background for the protests and their potential implications for US-South Korean relations (p. 403). The study provides evidence that the ideological affiliations of a news organization influences its coverage of foreign protest events.

2.4 THE UMBRELLA MOVEMENT

2.4.1 Background

The Umbrella Movement refers to a protest in Hong Kong that lasted from September 28 until December 15, 2014 that established encampments at several locations in Hong Kong. The movement emerged from the civil disobedience campaign “Occupy Central” which conducted a series of smaller protests in the first half of 2014. Originally, Occupy Central to pressure the leadership in Beijing to provide universal suffrage for the Chief Executive elections in 2017 (Lee and Chan 2018, p. 3).

Lee and Chan (2018, p. 13) trace to roots of the Umbrella Movement back to the July 1, 2003 protest, where Hong Kong citizens opposed the introduction of Article 23 and marks an important chapter in Hong Kong’s pro-democracy movement. With this first large-scale mobilization a “protest cycle” began (p. 25). The pro-democracy tendencies caused the Beijing government to exercise more intervention in Hong Kong, which in turn mobilized the pro-democracy camp (Ho 2019, p. 196). Since 2003, protesting became a new ritual and accepted norm for the general public in Hong Kong and they view protests as a legitimate channel to express political demands and give “diffuse support” to movements (Lee and Chan 2018, p. 26). The Umbrella Movement marks a “critical event” in this relationship with dialectic characteristics (Ho 2019, p. 196). It led to stronger decentralization and radicalization as it became evident in the latter half of 2019.

The idea of using civil disobedience as a tool for expressing political demands was introduced into the debate by Benny Tai via an essay in a Hong Kong periodical (Tai 2013; Lee and Chan 2018, p. 3). As a professor for law at the University of Hong Kong (HKU) his opinion carried weight in the public debate and he soon found two allies who would support his point of view. Together with

Chan Kin-man, a professor for sociology at the Chinese University of Hong Kong (CUHK), and Reverend Chu Yiu-ming he founded the *Occupy Central with Love and Peace* campaign (OCLP). They were often referred to as “The Trio” and their thoughts reached activists in Taiwan as well (Lee and Chan 2018, p. 3; Ho 2019, p. 120).

Two prominent student organizations adopted the ideas of *The Trio*, the long-established *Hong Kong Federation of Students* (HKFS) and newly founded *Scholarism*. Together with OCLP they initiated the occupy central campaign in the end of September 2014. Lee and Chan (2018, p. 78) argue that the Umbrella Movement was an “image event”, which describes the proliferation of strong imagery that turns something familiar into something strange. In this case it would be public spaces in Hong Kong being occupied by students and the police dispersing tear gas at them. The media broadcasting of the events of September 28th created the term “Umbrella Movement” as it featured the strong imagery of protesters opening their yellow umbrellas to defend themselves against the police usage of pepper spray and tear gas (p. 78). Surveys suggest that the protesters were motivated to join after they have seen the tear gas images on TV (p. 92).

Due to communication issues and the unexpected use of force by the police, the Umbrella Movement developed decentralized dynamics (p. 5). The Trio, HKFS and *Scholarism* were seen by the international media as the leaders of the movement, even though the protest participants did not necessarily share this view (p. 5). In fact, these three main organizers did not take a central role within the movement’s social network, which indicates the self-reinforcing characteristics of the movement (p. 139).

Ho (2019, p. 97) describes the concept of “movement-and-government standoffs” which defines “unusual episodes” in movements where the activists receive “focal attention of incumbents”. The activists are in a position where they can bargain over a longer period of time and also create a “possibility for a high-risk outcome” for either the activists or the incumbents (p. 97). The high stakes of these standoffs make well covered headline stories, as it was evident in the 1989 Tiananmen Movement or the 2011 protests in Egypt (p. 98)

Ho (2019) further suggests that the concept of threat has adequate explanatory power in regard to the rapid increase of movement participation and support in Hong Kong. The police using tear gas in Hong Kong against students established a threat for the citizens and in turn “exerted pro-mobilization power via strong emotional reactions” (p. 128). Lee and Chan (2018, p. 94) support this statement and suggest that the media played a role in spreading the information and the corresponding images that would evoke such emotional responses, which would render the media (especially TV) as an effective mobilization agent. Ho (2019, p. 124) further suggests that the police in Hong Kong immediately used tear gas to disperse the movement in order to prevent a similar occupation as it had occurred in Taiwan.

2.4.2 *Characteristics of the Umbrella Movement*

The majority of participants' age was under 25 with a post-materialistic value orientation (Lee and Chan 2018, p. 46). The participants had a tendency for an indigenous (Hong Kong) identity (Ho 2019, p. 60). Before the protests began, various groups such as *Occupy Central with Love and Peace*, *Scholarism*, and the Hong Kong Federation Students coordinated the campaigns. The ideological roots were in *Anti-Express Rail Movement* and a chain of previous campaigns (p. 86). They aimed for a specific policy goal, that was to revert the National People's Congress decision on candidate selection and universal suffrage. They pursued this goal with civil disobedience as core idea (Lee and Chan 2018, p. 52). That means that their speeches and texts emphasize non-violence and peacefulness. The movement was partially leaderless and became increasingly decentralized throughout the occupation (Ho 2019, p. 139-142). This decentralized approach also caused them to improvise coordination through digital media (Lee and Chan 2018, p. 182). But the three leading groups only partially represented the movement and thus were not able to coordinate well. This crippled negotiation leverage with the Hong Kong government (p. 187). And after approximately 50 days, the protesters became increasingly unsatisfied with *Scholarism* and HKFS leaders' speeches on the stage at the protest site (Ho 2019, p. 142). This dissatisfaction led to a diffuse and internal debate of about the core values and means of the movement. While some advocated that the movement is ought to remain peaceful and rational, other voices promoted ideas of violence and disruption (Lee and Chan 2018, p. 59). Lee and Chan (2018, p. 189) summarize that the protesters had a 'lack of a powerful collective action frame'. Throughout the occupation, the public support hovered at around 25-30% with a gentle decline (p. 150). Eventually, the protesters were forcefully evicted and were not successful in achieving its policy demands (p. 178).

2.4.3 *Umbrella Movement in previous content analysis*

Shahin et al. (2016) investigated the protest paradigm across Brazil, China and India by comparing the domestic coverage of selected protests in the three countries. Through a comparison of three samples of 360 articles from three countries, they can derive whether the protest paradigm can be transferred to non-Western media systems as well. Thus, they selected one conservative and one liberal newspaper for each country and chose a sample of 60 articles per outlet over a 30-day period where larger protests occurred in each country. The Brazilian sample was taken from June 2013, and the Indian one from April 2011 (p. 152). For China, they chose to investigate the Occupy Central campaign and randomly selected articles from two Hong Kong based newspapers, *Wen Wei Po* and *Apple Daily*. The sample period starts from September 26, 2014 the beginning of the protests (p. 152). The study considered nine variables that operationalize the protest paradigm and provide a rather fine-grained insight. The variables include sources, eyewitness account, public opinion, cause, cause evaluation, circus, appearance, peacefulness, violence, violence blame, protester evaluation, and administration evaluation (p. 153). Most of

the variables overlap with frames that were investigated before, but special attention is paid to the newspapers' potential to ridicule the protests by coding mentions of the protesters appearance, the presentation of the protest as a circus or spectacle, and occurrences where the cause was evaluated (Shahin et al. 2016, p. 151). The results indicate that the news outlets' ideological affiliations have a significant effect on the protest paradigm (p. 153). The data suggests that the coverage in *Apple Daily* and *Wen Wei Po* are almost exactly opposed to each other. The former relied on protesters for quotes, evaluated the cause positively, blamed the administration for episodes of violence as well as evaluated the government negatively (p. 155). The conservative paper on the other hand presented a negative evaluation of the protests and had a neutral stance on the government (p. 155). The Indian news outlets conformed the least with the protest paradigm, regardless of political orientation, which Shahin et al. (2016, p. 158) explain by the more informal political culture as well as the tendency of Indian media to be very critical of the government (p. 158).

Sparks (2015) compared the coverage of the Umbrella Movement in ten daily newspapers from the UK. He collected 156 articles, and analyzed their meta information such as length, date and authorship (p. 432). Further, he conducted a qualitative analysis of a few longer pieces and investigated their most common frames, tonality and issue salience (p. 431). The reporting most commonly framed the events in relation to the UK's historic and contemporary responsibilities towards Hong Kong, as well as in terms of economic consequences (p. 436). The protesters' cause was mentioned in almost all articles and their behaviors were presented as generally peaceful, while the police force was framed as aggressor (p. 438-439). Overall, the reporting was event driven and focused on particularly dramatic ones (p. 431).

Feng (2017) manually analyzed 120 news reports of four Hong Kong based Chinese language newspapers covering *Occupy Central*, namely *Ming Pao*, *Apple Daily*, *Oriental Daily News*, and *Ta Kung Pao*. The sample includes newspapers of all major political camps and ranges from August 31, 2014, the time when Beijing's announced the new electoral framework until September 31, 2014 when the occupation was fully deployed (p. 556). The framing of occupy central was evaluated through investigating the choices of presented attitudes in the text (p. 555). Journalists avoid expressing their opinion directly but would rather present a certain side of the story more prominently through quotes than opposing perspectives (p. 553). These statements potentially contain attitudes which implicitly or explicitly express appreciation of the event, the judgement of the occupiers, and the affect of the citizens (p. 563). The combination of explicit and implicit with the three aspects of attitude accounted for a total of six variables with levels for positive, negative and neutral (p. 555). The unit of analysis for this study were concordance lines around the keywords¹ "occupy central" or "occupy", meaning that the keywords are shown in context of a set amount of preceding words and following words. This created 856 concordances for manual analysis, but without stating the exact parameters of the

¹For the study Chinese search terms were used: "occupy central" (佔中) and "occupy" (佔領)

sliding window (Feng 2017, p. 556). The data of the study suggests that the papers supporting the movement (*Apple Daily* and *Ming Pao*) emphasized the justification of occupy central and the moral qualities of the protesters, while the pro-government newspapers stressed the public nuisances caused by the occupation, the protesters “selfishness”, and citizen statements expressing anger (p. 563).

Lee and Chan (2018) conducted the largest study on the Umbrella Movement which focused on the interactions between protests participants, traditional local media and social media. They conducted several on-site surveys, general opinion surveys, and content analyzed public posts on *Facebook*, as well as several Hong Kong based and Chinese speaking newspapers. A manual content analysis of 103 editorials and opinion articles from nine newspapers (*Ta Kung Pao*, *Oriental Daily*, *Apple Daily*, *Ming Pao*, *Sing Tao Daily*, *HK Economic Times*, *HK Economic Journal*, *HK Commercial Daily*, *Wen Wei Po*) ranging from September 1 to December 15, 2014 was conducted to evaluate the framing of the protest movement (p. 152). Lee and Chan (2018, p. 152) were specifically investigating three inductively derived counter-frames against the Umbrella Movement. The most prominent and oldest frame is the idea of foreign intervention in local affairs (p. 153). Where the protest is deemed to have its origin from foreign ideas or is even actively supported by other governments to undermine the stability of the society (p. 153). The frame of foreign powers meddling in Hong Kong’s affairs reaches back as far as the end of the 1980s and was introduced by the CCP and especially targeting the pro-democracy movement (Loh 2010, p. 175). It was also part of the critically rejected Article 23 reform in 2003 and was used in the latter half of 2019 as well. The second frame that appeared was the emphasis of rule of law in Hong Kong (Lee and Chan 2018, p. 159). It targeted the core idea of civil disobedience and spins the liberal progressive concept to favor a conservative audience (p. 159). The final and most simplistic frame was the emphasis on public nuisance, that carried weight in Hong Kong where major traffic routes were blocked by the protest camps (p. 161). Among the liberal newspapers in Hong Kong, *Apple Daily* and *Ming Pao* were both generally supportive of the movement and the frames were hardly present, especially in the early phase (p. 70). But the remaining mainstream media (including TV broadcast) conveyed a tonality bias against the movement (p. 195). The three counter-frames were frequently found in pro-government newspapers (*Oriental Daily*, *Wen Wei Po*, *Hong Kong Commercial Daily*) which conforms with the protest paradigm (p. 168).

2.5 SUMMARY OF DISCUSSED RESEARCH

Research on the media coverage of protests is rather diverse. Not all studies explicitly present their variables as frames but can be understood as such. Theoretical considerations across studies converge only partially, some studies refer to the protest paradigm intentionally, most can be interpreted with it in mind.

Almost all discussed works analyzed the presentation of protests as explicitly violent or peaceful, often as polar opposites. The inclusion of quotations in

the coverage is also a very common variable and it often includes the type of actor being quoted: domestic government or officials, protesters, bystanders, and foreign governments. Another researched variable is concerned with the sources of information used for the news item. These frequently included news agencies, official statements, protester statements, or even social media. Traditional newspapers are most likely to refer to government statements and official sources when covering foreign events, but also tend to include additional sources as well (Harlow and Johnson 2011; Karyotakis et al. 2017; Ha and Shin 2019; Chiao and Huang 2017; Watanabe 2017a; Du 2015). Especially, coverage of the Arab Spring in Chinese media outlets fits the pattern of the protest paradigm reasonably well (Chiao and Huang 2017; Du 2015).

Framing of the protest's cause is investigated in different ways. The simplest variable is whether the cause of the protesters is mentioned, and in which level of detail. It is often mentioned when it relates to foreign events along with background information, most prominently the *New York Times* (Harlow and Johnson 2011; Chiao and Huang 2017). Another common angle is to investigate who is being "blamed" or held responsible for the protest. That often includes government neglect, economic issues, protesters themselves, and even foreign intervention as origin for the protests. The imperialism and social stability frame by Ha and Shin (2019) are reminiscent of the foreign influence frame and rule of law frame by Lee and Chan (2018, p. 153). Especially this type of framing is frequently found in *China Daily*, *Al Jazeera*, TASS, or in pro-government newspapers in Hong Kong (Karyotakis et al. 2017; Ha and Shin 2019; Watanabe 2017a; Lee and Chan 2018). The visibility of protests in the daily coverage is a variable that is comparatively rare, but not to be neglected.

Ideological congruence appears to influence whether a newspaper conforms with the protest paradigm, regardless of covering domestic or foreign events (Du 2015; Chiao and Huang 2017; Karyotakis et al. 2017; Ha and Shin 2019; Lee and Chan 2018; Kim and Shahin 2020). However, journalistic traditions come into play as well, as it becomes evident in India, where there is a strong tradition to be critical of government action (Shahin et al. 2016). Among the reviewed works, only newspapers from the UK and South Korea contextualize foreign protest events with domestic issues (Ha and Shin 2016; Sparks 2015). Either this kind of "domestication" as described by Ha and Shin (2016, p. 538) is a rather rare phenomenon or it is not yet well understood.

Finally, most studies established an external frame of reference (Schröder 2002, p. 101) by defining events and comparing them against the results of the content analysis (Du 2015; Karyotakis et al. 2017; Lee and Chan 2018; Kim and Shahin 2020). Only the study by Lee and Chan (2018) compared the results of the content analysis with surveys as an extended frame of reference that enabled a search for potential media effects on the Umbrella Movement and public opinion (p. 19).

3 ANALYTICAL FRAMEWORK

Based on the literature reviewed, I propose an analytical framework for comparing the news media coverage of the protests in Hong Kong in 2014. The core of the comparison is based on an automated content analysis of selected English language newspapers from East Asia. Additionally, survey responses from the Asian Barometer Survey Wave 3 and Wave 4 are used to provide a small frame of reference to public opinion. In the first part of this section, the corpus and the sample criteria are defined. Next, I will present the criteria for comparison and how they are measured. Then I will describe the data collection and pre-processing steps. And finally, I will introduce how the criteria are tested for validity.

3.1 SAMPLE DEFINITION

The unit of analysis for this study are whole news articles including their headlines and teaser texts. Two English-language online newspapers for Hong Kong, Taiwan, China, and South Korea, are selected for sampling. Some countries have several online newspapers to offer, therefore, the top two with the highest website traffic are chosen by comparing their website ranking on *Site Rank Data* and *Alexa Internet*¹. As a result, the following newspapers are included in the sample: *South China Morning Post* (SCMP), *The Standard*, *Taipei Times*, *The China Post*, *China Daily*, *People's Daily*, *Chosun Ilbo*, and *The Korea Times*. Additionally, *The Guardian* is included in the analysis as a reference text that is from a non-Asian country. Comparing all UK based “quality” papers, *The Guardian* featured a relatively high visibility for the events in Hong Kong and is also a newspaper with a reasonably high circulation (Sparks 2015, p. 433). All criteria described below are also measured in the sample from *The Guardian* to serve as a baseline for comparison. It also enlarges the corpus, which is typically beneficial for the performance of the applied text models. The sample boundaries are from January 1 until December 31, 2014. This period covers not only the days of protest but also the early public discussions about occupying central up until the protest camps were forcibly removed (Ho 2019, p. 121).

The selected newspapers also exhibit a certain degree of variety in terms of their alleged partisanship. For Hong Kong, SCMP was considered skeptical towards China and mainland integration while maintaining a conservative standpoint that is favorable of the United Kingdom (Chan 2002, p. 103; *The New*

¹<https://siterankdata.com/> and <https://www.alexa.com/>; both visited on October, 22nd 2020.

York Times, [March 31, 2018](#)). On the other hand, *The Standard* is more tabloid-like and considered to be critical of the UK (Chan [2002](#), p. 102). The reports in *The Standard* are also more focused on local and “small” stories (p. 102). The *Taipei Times* is part of the *Liberty Times* which has close links to the Democratic Progressive Party (DPP, largest opposition party in 2014), and generally takes a “pan-green” stance on issues which includes skepticism towards China (Rawnsley and Rawnsley [2004](#), p. 6). *The China Post* allegedly has a “pan-blue” orientation which means that the paper’s political views align with the Kuomintang (KMT, ruling party in 2014) and is supposedly favorable of cooperation between China and Taiwan (Sullivan et al. [2018](#), p. 114). The *Chosun Ilbo* is one of the largest publications in South Korea and is considered to have a conservative orientation (J.-w. Lee [2011](#), p. 1). *The Korea Times* is also considered to be conservative, however, slightly closer to the center than the *Chosun Ilbo* (Molen [2011](#), p. 68; Antoniadis [2012](#), p. 12). Regarding the Chinese newspapers, their variety is very limited. The *China Daily* is a government sponsored newspaper with the ambition to provide a Chinese perspective on world issues (Rawnsley [2015a](#), p. 465). Similarly, the *People’s Daily* is owned and published by the CCP and its official mouthpiece (H.-w. Lee [2015](#), p. 122-123; Rawnsley [2015a](#), p. 465).

The expected variety in the sample, does of course not cover the whole spectrum of opinions that each countries’ news media landscape has to offer. Moreover, it is also not sufficient to portray the potential diversity in opinions across the populations. Nevertheless, the selected news outlets have combined a decent coverage and thus should provide a reasonable window into the reporting patterns.

In order to provide additional insights into public opinion, the results of the *Asian Barometer Survey* are used as an external frame of reference. Wave 3 (2010-2012) and Wave 4 (2014-2016) are selected, because they were conducted before the Umbrella Movement and after. This also allows to gauge whether the survey results are consistent over time or changing. The ABS results do not constitute the center for this thesis, but rather add another piece of information that helps drawing a better picture. The survey comes with imperfections, because items relating to protests are very limited. Nevertheless, it might be useful in assessing to which degree the newspapers are in alignment with public opinion.

3.2 CRITERIA

3.2.1 Overview

The criteria for comparison are based on previous works that investigated transnational coverage of protests in news media: salience, framing, quotations and sources. Each criterion is comprised of a set of indicators that are measured on the document level of the corpus and eventually aggregated the news outlet level. A summary of all criteria and their indicators is presented in Table [3.1](#). Based on the indicators, each criterion is evaluated according to its bias towards the Umbrella Movement. Thus, each criterion is ranked on a five-point scale from strongly in favor, over neutral, to strongly against the protests.

Table 3.1: Summary of criteria for comparison

Criteria / Indicators	Description	Technique	Document Level	Aggregated
Salience				
Coverage	proportion of documents mention Umbrella Movement	search string	dichotomous	continuous, 0–1
Overall Document Length	average length of all documents	word count	discrete	continuous
Relevant Document Length	average length of documents that mention Umbrella Movement	word count	discrete	continuous
Framing				
<i>only documents that mention Umbrella Movement</i>				
News perspective	average tonality	sentiment dictionary	continuous, -1–1	continuous, -1–1
Violent vs peaceful	average polarity score for word usage relating to “violent” or “peaceful”	LSS, standardized polarity score	continuous	continuous
Cause	proportion of documents mentioning protesters’ cause	keyword LDA, topic estimation	dichotomous	continuous, 0–1
Foreign influence	proportion of documents mentioning foreign influence on protesters	keyword LDA, topic estimation	dichotomous	continuous, 0–1
Quotations				
<i>only documents that mention Umbrella Movement</i>				
	proportion of documents containing quotes that can be attributed to 9 different actor categories	CoreNLP, quotation attribution	set of 9 dichotomous	set of 9 continuous, 0–1
Bought copy				
<i>only documents that mention Umbrella Movement</i>				
	proportion of news stories from external providers	search term dictionary	set of 14 dichotomous	set of 14 continuous, 0–1
<i>External Reference Criteria</i>	<i>Description</i>		<i>Respondent Level</i>	<i>Country Level</i>
Public Attitudes towards Protests				
<i>Asian Barometer Survey results, country level</i>				
Freedom to Protest	Proportion of respondents stating that the right to protest is essential in a democracy		dichotomous	continuous, 0–1
Freedom and Liberty	Aggregated mean score for respondents choosing answers relating to freedom and liberty		discrete, 0–4	continuous, 0–4

3.2.2 *Salience*

Editors of a newspaper face the decision on which stories to publish every day, and they select based on their evaluation of news value (Boumans and Trilling 2015, p. 11). Foreign events often carry less news value than domestic ones, thus it should not be taken for granted that foreign protests are reported in the news at all. If the editors decide to include news items about a foreign event, they also have to determine the length for the story. On the one hand, salience constitutes an important baseline in this study. On the other hand, international attention and visibility is generally a scarce resource that is assumed to be favorable for the protesters. As they become more visible by foreign audiences, their bargaining power also tends to increase. It also might shield them from drastic acts of violence by government actors. The salience is evaluated through two variables: relative occurrence and average document length. The identification of documents that mention the Umbrella Movement is also important because they constitute the “population of interest” for the other criteria (Krippendorff 2004, p. 122). The results are used for sub-setting corpus into a smaller population constitute the sample of relevant documents.

Coverage A Boolean search string is used to determine which documents mention the Umbrella Movement, Occupy Central, or Protests in Hong Kong. The search strings are based on the works of Lee and Chan (2018, p. 49) and implemented with the tool provided by Welbers, Atteveldt, and Benoit (2017) which allows for word proximity searches:

```
"Umbrella Movement*"~3 OR "Hong Kong protest*"~10 OR "HK protest*"~10 OR "Hong Kong demonstrator*"~10 OR "HK demonstrator*"~10 OR "Hong Kong Occupy Central"~10 OR "HK Occupy Central"~10 OR "Occupy Mong Kok"~10 "Occupy Admiralty"~10 "Occupy Causeway Bay"~10
```

The asterisk (*) enables partial word matching which includes possible inflections of words in the search results. The tilde character (~) with a following integer value restricts the sub-term to a word distance of 10 between all keywords within quotation marks and thus increases the chance of a correct hit (p. 261). The length of 10 based on the length of an average teaser sentence in news articles (Grusky, Naaman, and Artzi 2018, p. 711). This specific implementation of word distance ignores all non-word tokens such as punctuation marks or numbers.

As a result, the primary indicator for coverage is the proportion of documents where the search string found a match. This allows to balance out a potential difference in the total amount of articles that are published by the respective newspapers.

Document length The length of a document is calculated by summing all word tokens and excluding punctuation marks, special characters (e.g. emoji) or numbers. The word count is then aggregated by calculating the arithmetic mean of all documents mentioning the Umbrella Movement as well as all documents on news outlet level. This enables to compare the average length

against the news outlets internal preference and against all other news outlets in the corpus.

3.2.3 Framing

The variables for framing are only measured on the subset of documents that mention the Umbrella Movement. The variables for this criterion are designed that high and positive values indicate a favorable perspective on the protesters. Meaning that the tonality is overall positive, descriptions of peaceful protests are prominent, the cause is frequently mentioned and there is little mention about a foreign influence. Low and negative values indicate that the coverage is not in favor of the protests.

News perspective The overall perspective of a news item on the mentioned issues can be surrounded by a positive or negative tonality (Watanabe 2017a, p. 226). Tonality bias can be operationalized as counting the relative occurrence of words that carry positive or negative valence and thus yield a polarity score that typically ranges from -1 to 1 (Heidenreich et al. 2019, p. 1263). It is to be expected that news stories covering protest events will yield negative scores, because they inherently use words that carry a negative valence due to conflictual nature of the events (Watanabe 2017a, p. 226).

Many automated techniques for polarity analysis are based on search term dictionaries mentioned above. These dictionaries contain a long list of words, where each entry has a polarity score (Heidenreich et al. 2019, p. 1263).

For example, this sentence would have a negative polarity: “The government is *bad*”, while this example would carry a positive one: “I *love* the government”. The highlighted words are assigned with a polarity score that is weighted against the total amount of words in a sentence (Rinker 2019). There are several polarity (and sentiment) dictionaries for the English language available and the most sophisticated one can even account for words that shift the meaning and reaches an accuracy comparable with machine learning approaches (Naldi 2019, p. 6). Typical examples for valence shifters are words such as “not” or “very” as in: “The government is *not* bad” or “The situation is *very* bad”. In the first example, the meaning is reversed, shifting a word with negative polarity to a more positive one. The second example amplifies the negative polarity of “bad”.

The tool provided by Rinker (2019) has been validated and shows overall good performance on news content for tonality bias. The polarity scores are continuous values that range from -1 to 1 and are calculated on the document level. For the comparison the scores are aggregated on the news outlet level.

Violence vs peaceful One underlying value of Occupy Central was civil disobedience alongside non-violent protests (Lee and Chan 2018, p. 61). However, in the course of the protests there were episodes of violence by protest participants (Ho 2019, p. 122; Lee and Chan 2018, p. 75). Thus, it should be observable whether a news outlet emphasizes peaceful aspects of the protests or rather stresses aspects of violence. Descriptions of violence and peacefulness can be

understood as a concept of opposites or antonyms. Therefore, *Latent Semantic Scaling* (LSS) appears to be a suitable text model for measuring this type of valenced framing. LSS uses a small set of polarity seed words and “estimates semantic proximity” to other words in the corpus (Watanabe 2020, p. 6). For fitting the model, an entry keyword is recommended that is central to the analyzed concept. In this application, “protest²” is selected and the text model estimates terms that are commonly associated with the initial keyword by using a 10 word long sliding window. In the next step the polarity scores of the resulting words are calculated by their contextual similarity of the polarity seed words. This fitting process yields a dictionary of terms that are associated with the central key term together with a polarity value. This fitted dictionary can then be used to calculate the prevalence of these polarity terms on the document level and weight it against the total document length. The current implementation of the model automatically normalizes the results by centering them around the global mean and also scales it by a standard deviation (p. 8). The resulting polarity score for each document is then aggregated on the newspaper level.

Because LSS has not been employed in this specific use case before, the seed words are chosen with the help of the *New Oxford Thesaurus of English* and the results provided by Du (2015, p. 111). The seed words for peaceful are: “peaceful”, “harmonious”, “nonviolent”, and “pacifistic”. On the opposite end the chosen seed words are: “violent”, “vandalism”, “riot”, and “forceful”.

Additionally, Watanabe (2017a, p. 229) recommends using a separate corpus for fitting the model. The corpus should have the same text type (i.e. news stories) and should be considerably large. For this purpose, a sample of the *Newsroom* corpus appears to be a suitable choice. The corpus contains over 1.3 million English news articles from 1998 until 2017 (Grusky, Naaman, and Artzi 2018, p. 708). The corpus is first sliced to exclude documents after 2013, and then filtered to only include documents which mention include keywords related to protests². This size should be sufficient to ensure a decent amount of journalistic texts about protests.

Cause The framing of protests is strongly influenced by stating participants’ cause. At the extreme ends, newspaper editors may choose to cover the reasoning behind the protests in detail, or not even mention the concerns of the protesters at all (Lee and Chan 2018, p. 31; Kim and Shahin 2020, p. 402). A topic modelling approach is used to assess this frame. *Latent Dirichlet Allocation* is a probabilistic estimation that aims to unveil a “hidden” thematic structure in the corpus (Blei 2012, p. 79). The model takes document term matrix as input, which is a large matrix of *all* terms in the corpus against *all* documents. Based on the distribution of common term co-occurrences, the model estimates their relationship in comparison to other terms (Blei 2012, p. 79; Boumans and Trilling 2015, p. 15). The generative process requires a set of hyper-parameters where the most crucial is the number of topics to be estimated. Given that it is

²protest, riot, demonstrator, activist, marcher, extremist, insurgent, insurrectionist, radical, and rebel

difficult to know the number of topics a-priori, Maier et al. (2018, p. 110) suggest a series of steps to iteratively fit the model for optimal results, which is described below.

Once the model is fit to the corpus, it yields normalized probability values ranging from 0 to 1 for each term to be associated with each topic, and probability values for each document to be associated with each topic (Blei 2012, p. 80). Therefore, every document can be associated with several topics. The generated topics are collections of words that are likely to co-occur, and the researcher has to infer the topics based on the vocabulary. The original topic model by Blei (2012, p. 78) is initialized at random, which means that all terms are randomly assigned to the pre-set topics, and then iteratively re-distributed. However, as Maier et al. (2018, p. 103) points out, this leads to inconsistent results and thus it has proven useful to initialize non-randomly. While Maier et al. (2018, p. 104) suggest initializing the model based on term co-occurrence statistics, Eshima, Imai, and Sasaki (2020, p. 4) introduced a technique that allows to select a small set of keywords per topic for initialization. While the original LDA is a purely inductive approach, the keyword initialized version is a deductive-inductive one. Not all topics need to be pre-assigned with keywords, which enables the researcher to configure the model based on theoretical assumptions and fill the remaining parts based on the distribution of the data (p. 3).

In this study, it is of interest to determine whether the cause of the protesters is stated in the article. Therefore, one topic is initialized with the keywords: pro-democracy, universal, suffrage, democracy, election, 2017. The documents covering said topic are dichotomized based on a probability threshold of 0.5 (Maier et al. 2018, p. 108), mapping the variable to take values for framing being present or not. In a final step, the variable is aggregated on a news outlet level to a variable that indicates the proportion of documents from 0 to 1.

Foreign influence One common frame that was found in previous studies presents the protests as influenced by foreign entities. The notion that the protesters act upon the command of foreign powers undermines its legitimacy and makes it condemnable (Lee and Chan 2018, p. 153). This frame is also coded with the keyword assisted LDA and the topic is initialized with foreign, influence, conspiracy, conspire, sovereignty. Analogous to the cause frame, the topics in the 90th percentile are aggregated. This yields a final percentage of documents that employ the framing.

3.2.4 Quotations

News stories that conform to the protest paradigm are expected to rely on government sources for information and tend to neglect the protesters point of view. Accordingly, there is an expected tendency to quote government officials, police, or bystanders. Quote extraction and attribution can be determined by using the text model introduced by Muzny et al. (2017). It is a two-step approach build on top of the dependency parsing pipeline *CoreNLP* (Chen and Manning 2014). The dependency parser classifies the grammatical function of

each word of a given sentence and is also capable of recognizing named entities such as persons, countries, and organizations. The quote attribution model uses this information to detect quotes and to identify the speaker of a sentence (Muzny et al. 2017, p. 464). This yields a list of quotes per document together with the recognized entity. The model has been validated against canonically accepted tests and performed with a precision of 90.4% and recall values of 65.5% (p. 464). These numbers suggest that the model makes conservative predictions by missing about one third of all quotes and falsely attributes 1 out of 10 quotes.

This list is then further processed, and each recognized entity is allocated to a category of actors. Based on the literature reviewed in section 2.4, the relevant actor groups are summarized as follows:

- Hong Kong officials: government representatives, (former) Chief Executives, and police officers
- Hong Kong Protest Camp: Occupy Central Leaders, Student Leaders, and anonymous protesters. The actor names and organizations that are associated with the protest camp are curated based on the works by Lee and Chan (2018, p. 61-77) and Ho (2019, p. 63).
- Bystanders: “a woman”, “a man” and other unnamed persons.
- Chinese officials: all kind of government members, representatives or CCP members.
- Counter Protesters: representatives of the organized anti-occupy camp, e.g. “Alliance for Peace and Democracy”.
- Newspaper’s home government: all officials that can be linked to belong to the home country of the newspaper.
- Other governments: all officials that belong to a government which is neither one of the above, e.g. US President.
- Sunflower Movement: members of the Sunflower Movement in Taiwan.
- Unknown: quotes where the text model cannot provide attribution.

This process yields seven dichotomous variables per document, where 0 represents no quote from an actor group and 1 indicates at least one corresponding quote. The result is then aggregated on the news outlet level and thus provides an overview of the newspaper’s quotation preferences.

3.2.5 *In-house vs bought copy*

Newspapers frequently rely on the services of news agencies to provide a rich variety of news content. It is comparatively expensive to maintain offices in foreign countries in order to provide occasional reporting on foreign affairs. Therefore, it is to be expected the selected news outlets feature articles that they bought from news agencies or other news outlets. As the research by Ha and Shin (2019, p. 210) suggests, there are differences in the framing between articles authored by in-house journalist and those originating from press agencies. While the main focus of this analysis is not to make this type of comparison, this data point might help to identify potential patterns that different newspapers

share. In case some newspapers frequently rely on the same press agency for their reporting on the protests in Hong Kong, most of the criteria above should be similar. Therefore, to control for such patterns, the source of the document is assessed by using a search keyword dictionary that covers the following ten press agencies:

- *Agence France-Presse* (AFP)
- *Al Jazeera*
- Associated Press (AP)
- Bloomberg News
- Central News Agency (CNA, Taiwan)
- Kyodo News
- Reuters
- TASS
- Xinhua News Agency
- Yonhap News Agency

Additionally, the dictionary includes all nine newspapers and the following news outlets:

- Radio Television Hong Kong (RTHK)
- Global Times
- Voice of America (VOA)
- Others

It is assumed that a news story can have only one news agency as source. If there is no hit for an external source, it is assumed that the article has in-house authorship. The hits of the dictionary are aggregated on the news outlet level in ten variables. Each variable ranging from 0 to 1, indicating the proportion of articles that were identified from a respective news story.

3.2.6 *Public Attitudes Towards Political Protests*

There are no internationally coordinated surveys available that specifically asked questions relating to the Umbrella Movement. Therefore, for the sake of this thesis, only general attitudes towards protests are taken as a proxy value. The ABS Waves 3 and 4 both include question items for political participation and action. However, this item does not seem to be a suitable proxy, as the personal participation behaviors do not automatically imply agreement or disagreement with other citizens' political actions. Additionally, only a very small minority of survey respondents stated that they had participated in a protest before.³

A more suitable proxy appears to be the respondents' views on political rights, where the ABS also asks about the right to participate in protests. This

³In Wave 3 and 4 only less than 5% of respondents stated that they had participated in a protest at least once. (Asian Barometer Survey, [Wave 3](#); [Wave 4](#)).

question is included based on the experience that East Asians in general, regardless of degree of political rights in their respective countries, would describe their government as democratic (Chu and Huang 2010, p. 115). Therefore, starting from ABS Wave 3 question batteries that dig deeper in this issue are included. Thus, Wave 3 and Wave 4 feature a survey battery comprised of four questions that gauges the respondents' personal understandings of democracy (Lu 2012, p. 4). One of the questions includes an item that is concerned with protests. The questions are designed to cover four "components in popular democratic conceptions" and include 1) norms and procedures, 2) freedom and liberty, 3) social equity, and 4) good government (p. 4). The questions are asked with the following leading statement:

"Many things may be desirable, but not all of them are essential characteristics of democracy. If you have to choose only one from each four sets of statements that I am going to read, which one would you choose as the most essential characteristics of a democracy?" (p. 5)

Then a series of four questions with four answer items is asked, where each answer possibility matches one of the four components of democracy and only a single choice is possible (p. 28):

Measure 1:

1. Government narrows the gap between the rich and the poor. (social equity)
2. People choose the government leaders in free and fair election. (norms and procedures)
3. Government does not waste any public money. (good government)
4. People are free to express their political views openly. (freedom and liberty)

Measure 2:

1. The legislature has oversight over the government. (norms and procedures)
2. Basic necessities, like food, clothes and shelter, are provided for all. (social equity)
3. People are free to organize political groups. (freedom and liberty)
4. Government provides people with quality public services. (good government)

Measure 3:

1. Government ensures law and order. (good government)
2. Media is free to criticize the things government does. (freedom and liberty)
3. Government ensures job opportunities for all. (social equity)
4. Multiple parties compete fairly in the election. (norms and procedures)

Measure 4:

1. People have the freedom to take part in protests and demonstrations. (freedom and liberty)
2. Politics is clean and free of corruption. (good government)
3. The court protects the ordinary people from the abuse of government power. (norms and procedures)
4. People receive state aid if they are unemployed. (social equity)

According to Lu (2012, p. 6), the answer possibilities can then be aggregated to construct a variable that represents the personal understanding of democracy. Each question is transformed into a set of dummy variables and then they are simply added up to result in an index from 0 to 4. Therefore, a respondent who chooses the answer relating to liberty and freedom for all four questions would receive the value 4 for this aspect and 0 for the other three.

For this study, the freedom and liberty aspects are particularly relevant, because it is assumed that it represents a certain ideological alignment between the survey respondents and the Umbrella Movement. As it was described above, ideological alignments appear to be relevant for the framing of social protests (Kim and Shahin 2020, p. 402; Ha and Shin 2016, p. 551). Therefore, the corresponding questions of the survey are used as criteria for comparison. The first criterion is the count of respondents who chose the answer relating to people's rights to participate in protests. The second is the freedom and liberty variable as suggested above. As hinted above, there are weaknesses to this approach. The personal understanding of political rights in a democracy might not exactly reflect personal attitudes towards protests. However, the respondents' personal prioritization for the right to protest can be understood as a sign that they value it as a means for political participation.

3.3 DATA

3.3.1 Data collection

The raw news stories are collected by systematically scraping the newspapers online editions:

- *South China Morning Post*: <https://www.scmp.com/>
- *The Standard*: <https://www.thestandard.com.hk/>
- *The China Post*: <https://chinapost.nownews.com/>
- *Taipei Times*: <https://taipeitimes.com/>
- *Chosun Ilbo*: <http://english.chosun.com/>
- *The Korea Times*: <https://www.koreatimes.co.kr/>
- *China Daily*: <http://www.chinadaily.com.cn/>
- *People's Daily*: <http://en.people.cn/>
- *The Guardian*: <https://open-platform.theguardian.com/> (access point for developers)

For this purpose, a collection of custom scripts is deployed which download every article that was published between the January 1 and December 31,

2014. Most newspapers do not offer a dedicated application programming interface (API) to enable bulk downloading of content. Therefore, the scripts are configured to browse the websites and download the whole hypertext markup language (HTML) page for later processing.

To minimize redundant or not relevant data the pages are filtered, and specific newspaper sections are excluded. For most websites, this step can be achieved during the scraping process, but some require manual filtering of the data after it has been downloaded. The excluded pages are articles that belong to sections related to sports, weather, entertainment (e.g. film, fashion, food, television), lifestyle, music, travel, or miscellaneous such as crossword puzzles. Arguably, this leaves a gap in the data, but for the purpose of this paper, it is sufficient to ensure that news stories relating to sections such as politics, finance, business, opinion, or foreign affairs are not filtered out.

3.3.2 Data Pre-processing

Before the raw data can be turned into a corpus, it has to undergo a series of pre-processing steps. The goal of this part is to harmonize the data and create consistency (Krippendorff 2004, p. 84). Aschenbrenner (2019, p. 146-148) and Maier et al. (2018, p. 101) provide a useful overview of required steps and highlight that correct pre-processing is crucial for later analysis. First, the HTML code is parsed for its structure and the core elements of the news article are extracted (section, publication date, headline, teaser text, and main text). Remaining HTML fragments are removed as well as e-mail addresses or internet links. The latter two are replaced with “EMAIL_ADDRESS” and “HYPERLINK”. Additionally, punctuation marks are normalized such as replacing typographic quotation marks (“ or ”) with a standardized character ("). In the next step the single documents are bundled into a table, where the columns include the text of the document, and meta data such as publication date and news outlet.

The next step is to divide the corpus according to articles that mention the Umbrella Movement as described above. The tool provided by Welbers, Atteveldt, and Benoit (2017) can handle large text data efficiently and slice it according to search terms. The following processing steps are only applied to the resulting subset.

For deeper analysis it is required to tokenize and lemmatize the corpus. Tokenization splits the text into single items (*tokens*) at every word boundary which are mainly space characters and punctuation marks. Lemmatization identifies the dictionary form of each word and thus removes possible inflections of each word. Lemmatization improves the data quality and harmonizes it which is especially relevant for automated approaches (Aschenbrenner 2019, p. 147). The dependency parsing pipeline *CoreNLP* by Chen and Manning (2014) is capable of handling this task with a high accuracy for English. Additionally, it can also perform part-of-speech tagging (POS-tagging) which includes identification of word types, and their syntactic functions, and it can also recognize named entities and attribute quotations. The result of this step is a large table, where each row is a single token, and the columns hold infor-

mation such as document id, sentence id, and all of the parsing results as well as the meta data from the steps before. Once the corpus is in this condition, it is ready for most analysis techniques.

3.3.3 *Specific pre-processing for LDA*

Latent Dirichlet Allocation requires the corpus to be transformed into a document-term-matrix, where each row represents a document and each column a term. The values correspond to the frequency of the term in the document. A direct conversion from the corpus into the matrix would lead to non-optimal results, because there is a loss of semantic information that is important for later analysis (Maier et al. 2018, p. 101). For example, persons such as *Abe Shinzo* or *President Park* would be separated into two or more terms. Furthermore, compound words such as “tear gas” or “police force” would be separated as well. To minimize the information loss named entities and compounds are concatenated to single terms such as “President_Park” or “tear_gas”. Next, stop words are removed which are terms that do not “convey meaning but primarily serve grammatical functions” (Grimmer and Stewart 2013, p. 273).

After these steps are completed, the document-term-matrix is created, and the data can be further reduced. Maier et al. (2018, p. 101) highlights the importance of relative pruning, which means to reduce the vocabulary of the matrix by removing very frequent or very rare words. They recommend removing terms that “occurred in more than 99% or less than .5% of all documents” (p. 101).

As mentioned above, LDA requires to set the number of topics beforehand, but it is possible to run several models iteratively with a varying number of topics and tune for the optimal results. In this tuning process, the model is run against the majority of the data and then it is calculated how well it predicted the remaining data (Jacobi, Atteveldt, and Welbers 2015, p. 95; Maier et al. 2018, p. 101). As Jacobi, Atteveldt, and Welbers (2015, p. 94) suggest, a sensible solution is to optimize for *perplexity*. Perplexity indicates how well the model can predict the distribution of terms in the topics correctly (p. 95). Usually, a higher number of topics leads to a better prediction, but it also means that more topics lack human interpretability. The predictive quality improves asymptotic against the number of topics. Therefore, the researcher has to decide at which point the number of topics is sufficiently low without greatly increasing the predictive quality of the model (p. 95). There are also other similar indicators that are optimized for maximization or minimization of topics (e.g. Griffiths and Steyvers 2004; Cao et al. 2009). Fortunately, these indicators are combined in the tool provided by Murzintcev (2020), which allows to automatically run various models and gather the results. Thus, using these indicators, the optimal value for the number of topics is determined, before the keyword assisted topic model is run.

3.4 EVALUATION OF CONTENT ANALYSIS RESULTS

3.4.1 *Validation*

Automated text analysis carries a risk of generating misleading or wrong results, therefore validation serves to minimize this risk (Grimmer and Stewart 2013, p. 271). The validation method depends on the nature of the automated analysis. Search term-based methods are typically validated against a human coded sample set and tested for face validity (Krippendorff 2004, p. 319). In principle, it is assessed to which degree the human coded results and the search term hits are in agreement. Inductive approaches that are based on statistical distribution such as LDA require a more complicated approach for validation, because the results of the model are not known a-priori (Maier et al. 2018, p. 93).

Search string-based variables The variables generated for daily coverage, news perspective, violence, and news agency are all in principle based on search term hits. Therefore, the results of these variables are optimized to be similar with human generated results. For the purpose of this thesis, the variables are tested for either face validity or correlated validity of a small human coded sample set (Krippendorff 2004, p. 321; Watanabe 2017c, p. 4). Face validity tests are used for dichotomous variables where there the computer can be either right or wrong. Correlative tests are for variables that measure along a continuous scale.

Two different sample sets are created and manually annotated. Ideally, the sample sets would at least include 300 documents to achieve an adequate level of significance (Krippendorff 2004, p. 122). However, given the constraints of this thesis regarding costs and time smaller samples with a minimum of 30 are generated (Wimmer and Dominick 2011, p. 103). The first sample considers the entire corpus and is for validating the search strings that determine whether a document mentions the Umbrella Movement. It is to be expected that the Umbrella Movement appears in only a small number of documents and thus the probability of generating a random sample with sufficient hits is low. To improve on this, a partial bias is introduced into the random sample (Krippendorff 2004, p. 116). The total sample consists of 100 news articles. 50 are randomly sampled from the entire corpus, 25 from the first days of protests (28.09.2014–01.10.2014), and 25 from the last days of protests (11.12.2014–15.12.2014).

The second set is taken from the population of interest, which are documents that mention the Umbrella Movement. The random sample consists of 30 articles and is manually annotated for two variables:

- Violence framing: Coding on a five-point scale, where 1 equals violence framing is present, 3 that the article is balanced or has no such framing, and 5 a peaceful framing (Watanabe 2017c, p. 4).
- Sources: Coding of news agency according to the list provided above.

The variable for news perspective uses a library that was validated by Rinker (2019); therefore, no further validation is conducted.

LDA based variables Maier et al. (2018, p. 108) provide a series of recommendations for validating the results of LDA. Regardless, whether the model is initialized at random, based on co-occurrence statistics, or with keywords, the external validation procedure is the same. In this case, first the optimal number of topics for the corpus is determined, as described above. Then the model is initialized where some topics are assigned fixed keywords. The pre-defined topics are all based on results from previous studies:

- foreign influence (Lee and Chan 2018, p. 153): foreign, influence, conspiracy, conspire, sovereignty
- cause (Lee and Chan 2018, p. 31; Kim and Shahin 2020, p. 402): pro-democracy, universal, suffrage, democracy, election, 2017
- handover (Sparks 2015, p. 436): handover, former, colony, crown, 1997, declaration
- rule of law (Lee and Chan 2018, p. 159): illegal, law, order
- public nuisance (Lee and Chan 2018, p. 161; Feng 2017, p. 563): rubbish, trash, paralyze, traffic, livelihood, inconvenience, disruption

The purpose of explicitly including other frames as well is to prevent the model from combining topics that may co-occur frequently.

Once the number of topics has been determined and the keywords are chosen, there are three strategies for optimizing the interpretability of the topics. It is possible that some can be summarized based on their distribution and semantic criteria, or even to be discarded. LDA often yields a series of background topics that are present in a large number of documents but have an overall low probability of being associated with them. They “contribute much to the whole model, but their word distribution is not very specific” (Maier et al. 2018, p. 107). Therefore, they can be discarded from the further analysis. Another strategy is to sort the terms within a topic by their relevance, meaning that terms which rarely occur in the corpus and not in other topics are weighted higher. This causes more generic terms to rank lower and makes the topics more specific and thus improves their interpretability (p. 107). The third strategy is to include meta information in the topic distribution and look for their concentration. For example, the news outlet can act as a meta information and it can be calculated if the topics are equally distributed amongst all news outlet or specific to a few (p. 107).

Next, the topics are assessed by humans through a sample, where 10 random documents are selected for each topic where the prediction value is higher than 0.5 (p. 108). Based on the weighted word list associated with the topic and the sample documents, the human readers aim at eliminating topics of low semantic quality. In this step the validity of the keyword-initiated topics is also qualitatively assessed (Eshima, Imai, and Sasaki 2020, p. 18-19).

The final result of these steps is grouped a table of topics that shows their human assigned label, top terms, and prevalence in the corpus. The topics are grouped based on semantic categories that are derived from their similarity. The table should also indicate which topics were assessed with low semantic quality or are categorized as background topics (Maier et al. 2018, p. 109).

3.4.2 *Significance testing*

While the final criteria for comparison are based on the aggregated news outlet level, additional steps are taken to compare the variables on a document level. The differences in the distributions of the variables is tested for statistical significance by computing χ^2 -tests or performing one-sided ANOVA (Gunter 2002, p. 233; Backhaus et al. 2016, p. 175). The χ^2 -test is commonly used for hypothesis testing in a variety of content analysis studies that compared news outlets based on framing and other variables (Ha and Shin 2016, p. 545; 2019, p. 210; Lee and Chan 2018, p. 169). This test is suitable for comparing nominal or categorical data, which in this case is the news outlet of the document. If the test indicates that the result for one variable is not significant, it means that there is no difference between the news outlets and that they overall share the same reporting style for this variable. The test is conducted for the dichotomous and nominal variables: daily coverage, cause, foreign influence, quotations, and sources. The continuous variables for word count, news perspective and violence framing are tested via one-sided ANOVA tests (Backhaus et al. 2016, p. 175; Lee and Chan 2018, p. 55).

4 EMPIRICAL PART

In this section, I present the results of the data collection, processing, analysis, and validation. First, I briefly describe the collection of online articles, then how the relevant documents which mention the protests in Hong Kong were identified. The subsequent subsections are organized by the used content analysis technique. Therefore, the cause frame and foreign influence frame are presented together, because they rely on the same text model. The last subsection presents the relevant data from the *Asian Barometer Survey* Wave 3 and Wave 4.

4.1 COLLECTED TEXT DATA

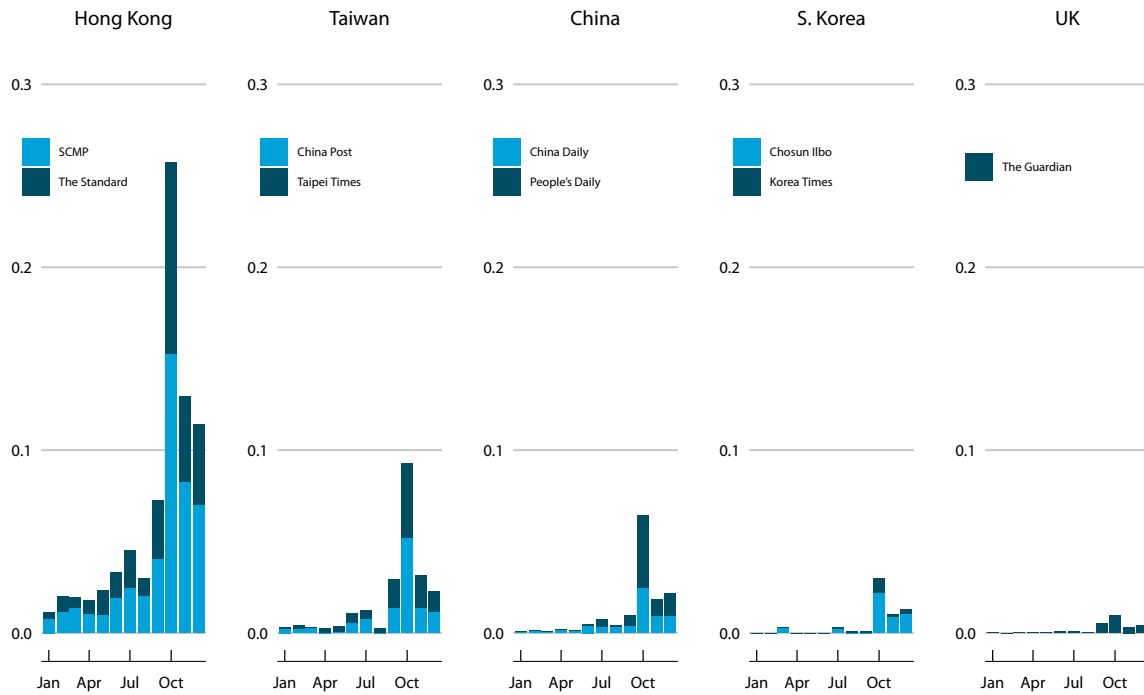
During the period of early February until late March 2020, a total of 221,640 unique documents of the nine news outlets were scraped. The highest yield is from *China Daily* with 66,276 documents and the lowest was collected from *Chosun Ilbo* with 4,334. Table 4.1 shows an overview of the collected data and the average amount of documents that the newspapers published daily.

Table 4.1: Overview of collected documents and search string hits

Territory	Newspaper	Documents	Daily average	Mentioned	
				<i>n</i>	%
China	China Daily	66,276	181.6	376	0.57
	People's Daily	9,619	26.4	76	0.79
Hong Kong	SCMP	24,273	66.5	911	3.75
	The Standard	16,098	44.1	381	2.37
S. Korea	Chosun Ilbo	4,334	11.9	17	0.39
	Korea Times	17,230	47.2	21	0.12
Taiwan	China Post	15,889	43.5	152	0.96
	Taipei Times	19,611	53.7	175	0.89
UK	The Guardian	48,310	132.4	88	0.18
<i>Total</i>		221,640	607.2	2,197	0.99

Column *Daily average* was calculated by dividing document count by 365. A χ^2 -test of independence was calculated to examine the relation between newspaper and search hits (column: *mentioned*). It shows that the newspapers differ from each other significantly: $\chi^2(8) = 2795.20, p < .001$. Difference between countries is significant as well: $\chi^2(4) = 2598.44, p < .001$.

Figure 4.1: Search hits results as proportion of monthly total documents. Grouped by country and outlet.



4.2 SALIENCE

4.2.1 Daily Coverage

The search string for identifying documents mentioning the Umbrella Movement or Occupy Central campaign generated 2,197 hits. As Table 4.1 shows, the *South China Morning Post* has the highest amount and the *Chosun Ilbo* the lowest with 911 and 17 respectively. Considering the proportion of hits against the total amount of documents per newspaper, SCMP has still ranks one, but *The Korea Times* has the lowest proportion with 0.12%.

The news coverage of protests is also subject to special events. Therefore, the proportional search hits were also calculated on a monthly level and plotted in Figure 4.1. The figure shows that October 2014 has the highest monthly proportion of search hit results.

The validation of the search string results reports a face validity of 0.92. Meaning that the agreement between human coding and search string hit is at 92%. The search strings did not miss any document, but rather selected too many. The manual inspection of search hits in the period before the protests started reveals that there are also a handful of documents that cover protests in Hong Kong concerned with different issues. For example, some documents mention a protest regarding poultry regulation (e.g. *Taipei Times*, [January 29, 2014](#)), others are covering a few episodes of foreign domestic workers protesting in Hong Kong:

Table 4.2: Average document length by Newspaper

Territory	Newspaper	Document Length (word count)					
		All		Not mentioned		Mentioned	
		Mean	sd	Mean	sd	Mean	sd
China	China Daily	397.3	370.2	396.2	370.3	602.4	271.4
	People's Daily	235.9	220.0	234.2	219.3	442.4	210.9
Hong Kong	SCMP	493.9	312.0	483.3	241.3	766.1	1,011.8
	The Standard	213.5	121.6	211.9	120.8	279.1	136.9
S. Korea	Chosun Ilbo	169.6	112.7	169.0	112.2	309.1	146.6
	Korea Times	372.6	272.9	372.3	272.6	633.5	341.7
Taiwan	China Post	348.6	127.5	348.2	127.5	395.0	120.2
	Taipei Times	458.4	204.6	457.6	203.8	546.8	263.9
UK	The Guardian	737.3	470.7	737.2	470.9	798.5	372.7
	<i>Overall</i>	457.2	376.5	455.8	371.6	595.8	700.3

N = 221,640. Mentioned = 2,197. Levene's test indicated unequal variances for the whole sample ($F = 1697.8, p < .001$) and also the subset of *mentioned* ($F = 8.35, p < .001$). The difference of document length is statistically significant for the different newspapers of the whole sample: Welch's $F(8, 51201) = 9783.7, p < .001$. Also for the subset of *mentioned*: Welch's $F(8, 182.75), p < .001$.

“Thousands of domestic helpers took to the streets of *Hong Kong* in *protest* in January demanding justice over the case. Her employer has been arrested and charged with assault.” (SCMP, [March 7, 2014](#), emphasis added)

Or also protests against Japan:

“When our *protest* in *Hong Kong* was over, I texted many friends, saying, ‘I just went to *protest* at Consulate General of the United States in *Hong Kong*,’ with only a dozen people, six media – five print and one electronic.” (China Daily, [May 12, 2014](#), emphasis added)

4.2.2 Document Length

The average document length in the corpus is at approximately 456 words and documents that mention the protests are on average longer as presented in Table 4.2. *The Guardian* has on average the longest documents in the sample regardless of mentioning the protests or not. The SCMP has on average the second longest documents mentioning the protest, but also the highest standard deviation. While the *Chosun Ilbo* has the shortest regular document length, *The Standard* has the shortest length in the mentioned category. A one-way Welch test comparing each newspaper indicates that the document lengths differ significantly when they mention the protests.

Table 4.3: Average tonality of documents mentioning the protests

Territory	Medium	Sentiment Score	
		Mean	sd
China	China Daily	-0.01	0.11
	People's Daily	-0.04	0.11
Hong Kong	SCMP	-0.01	0.10
	The Standard	-0.01	0.11
S. Korea	Chosun Ilbo	-0.03	0.10
	Korea Times	-0.02	0.11
Taiwan	China Post	-0.03	0.12
	Taipei Times	0.00	0.11
UK	The Guardian	0.00	0.08
Overall		-0.01	0.11

$N = 2,197$. Levene's test indicated unequal variances ($F = 2.38, p < .01$). The sentiment scores do not differ significantly between the newspapers: Welch's $F(8, 180.28) = 1.59, p = .13$

4.3 FRAMING

4.3.1 News Perspective

The analysis of tonality shows that *People's Daily* has on average the most negative score, meaning that they tend to use more words that the sentiment dictionary has indexed with a negative value. Overall, all newspapers tend to be on the negative side except the *Taipei Times* and *The Guardian* as shown in Table 4.3. However, considering that the protests featured episodes of violence and clashes with the police force and other groups, it is not surprising that the documents on average score negatively. A one-way Welch test shows that there is no significant difference between the newspapers, rendering the initial analysis of this criterion as inconclusive. Subsequent Welch Two Sample t-tests of newspaper pairs by country showed that the only statistically significant difference is found in Taiwan. The difference in mean sentiment scores of *The China Post* and *Taipei Times* is statistically significant, $t(314.39) = -2.38, p = .018, d = -0.26$.

4.3.2 Violent vs Peaceful

The *Newsroom* corpus was filtered to exclude articles from 2014 and also to only include articles that mentioned words relating to protests. From a theoretical perspective, the latter step is not necessary, but otherwise the corpus with 1.3 million articles would have been too large to fit in 32 GB of memory. After sub-setting, the training corpus contained 125,039 articles, which is more than were used by Watanabe (2017a, p. 229) and thus, sufficient. Fitting the LSS-text model with the newsroom corpus generated a dictionary with 6,343 terms. The terms associated with "peacefulness" received a positive value and terms related to violence a negative value. The top terms at both ends of the dictio-

Table 4.4: Top 20 positive and negative terms of dictionary generated with LSS

Top positive		Top negative	
Term	Score	Term	Score
<i>peaceful</i>	0.233	<i>vandalism</i>	-0.358
nonviolence	0.217	riot	-0.347
conciliatory	0.175	incidents	-0.312
interfaith	0.165	arson	-0.303
buddhists	0.160	<i>violent</i>	-0.299
ideals	0.159	explosion	-0.295
<i>nonviolent</i>	0.146	incident	-0.292
buddhist	0.144	<i>riots</i>	-0.286
reconciliation	0.142	police	-0.285
peace	0.136	assault	-0.284
principles	0.135	cops	-0.283
peacefully	0.135	storm	-0.282
secularist	0.124	caused	-0.278
ideal	0.123	rage	-0.278
philosophy	0.119	crime	-0.277
shafiq	0.118	shootings	-0.273
tibetan	0.115	warning	-0.268
traditions	0.114	victim	-0.268
orderly	0.112	blasts	-0.268
mediterranean	0.111	vandals	-0.268

Total dictionary length: 6,343 terms. Seed words are highlighted.

nary are presented in Table 4.4. The distribution of terms is shifted towards the negative end with a mean of -0.066 (median -0.067), meaning that there are more terms with negative values in the dictionary than positive ones. Most terms fit with the ideas of peacefulness and violence, however, there are certain terms that might not be directly related to protests. On the “peaceful” end the seed word “harmonious” probably pushed the model to include terms such as “Buddhist”, “interfaith” or similar terms related to religion. However, it is also plausible to assume that newspaper articles generally mention these terms in contexts related to peacefulness. On the end of “violence”, it appears that the terms “police” and “cops” often co-occur with terms related to violence. For the task at hand, it is possible that these types terms could introduce a bias into the analysis. Considering that there are documents covering press conferences of the police force, it is possible that these get classified incorrectly. To counteract this potential interference, the terms “police” and “cops” were removed from the dictionary.

Applying the dictionary to the population of interest in the corpus shows that the SCMP has on average the most negative score of -0.134 and the *People's Daily* the most positive score with 0.449, which is presented in Table 4.5. However, both means also have considerably high standard deviations of 0.975 and 1.046, respectively. As expected, the mean of the entire sample is exactly zero and the standard deviation is 1. An one-sided Welch test suggests that the

Table 4.5: Summary of results for fitted LSS Textmodel

Territory	Medium	LSS Score	
		Mean	sd
China	China Daily	0.372	0.913
	People's Daily	0.449	1.046
Hong Kong	SCMP	-0.134	0.975
	The Standard	-0.120	1.071
S. Korea	Chosun Ilbo	-0.044	1.285
	Korea Times	0.289	0.856
Taiwan	China Post	-0.082	1.063
	Taipei Times	0.011	0.932
UK	The Guardian	-0.016	0.718
Overall		0.000	1.000

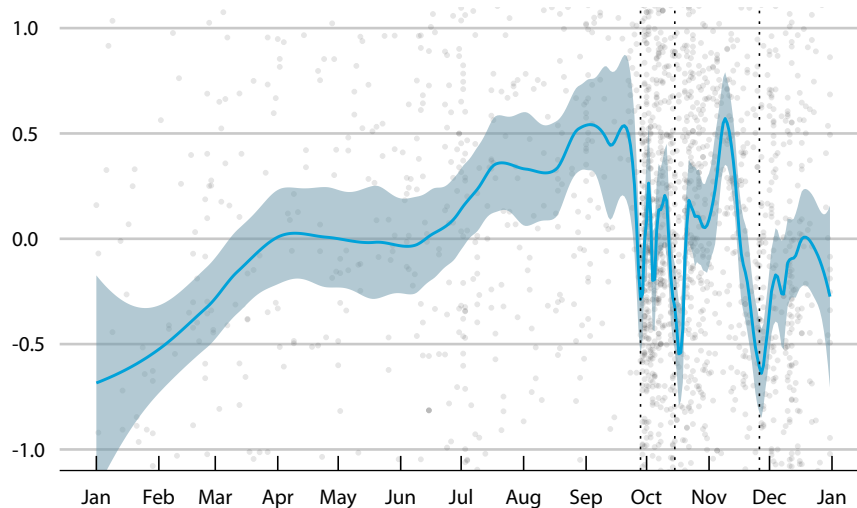
N = 2,197. Minimum: -4.36; maximum: 2.70; median: 0.026. Levene's test indicated unequal variances for medium comparison ($F = 2.77, p < .001$) and country comparison ($F = 2.98, p = .018$). The LSS scores differ significantly between the newspapers: Welch's $F(8, 180.2) = 12.37, p < .001$; and countries: Welch's $F(4, 200.5) = 24.48, p < .001$.

differences between the newspapers are significant.

To assess the external validity, the resulting LSS scores are plotted on a date axis in Figure 4.2. Considering that the protest period featured episodes of violence, it is to be expected that the overall scores on the days of reporting should be negative. As the figure indicates, there are three major dips, each corresponding to an extraordinary event. On September 28, starting from the evening until late at night, the police attempted to disburse the crowds and to prevent the protesters from establishing encampments, leaving a considerable number of protesters injured (Lee and Chan 2018, p. 77-78). Next, on the night from October 15th to 16th, district council member Ken Tsang who participated in the protests was singled out by police officers and severely beaten (p. 99). A TVB camera operator captured the incident and it aired the following morning (p. 99). The final dip is on November 26, the day when the police concluded the eviction of the Mong Kok protest site (p. 178). All three dips in Figure 4.2 are a day after an extraordinary event of violent behavior happened, which shows that the model appears to be generally able to cover the semantic spectrum of terms related to "violence" and "peacefulness".

Subsequent post-hoc tests suggest that not all newspapers differ from each other significantly. Two Games-Howell tests (Hilton and Armstrong 2006, p. 36) were performed for comparing the LSS scores on newspaper and country level. *China Daily* shows the most significant differences with several other newspapers: *The China Post*, *SCMP*, *The Standard*, *Taipei Times*, and *The Guardian* ($p < .05$, see Appendix for full table). In the listed comparisons, *China Daily* has a more positive score than the other newspapers. Similar results appear for *People's Daily*, however, it only differs significantly with

Figure 4.2: Latent semantic scaling results as longitudinal plot.

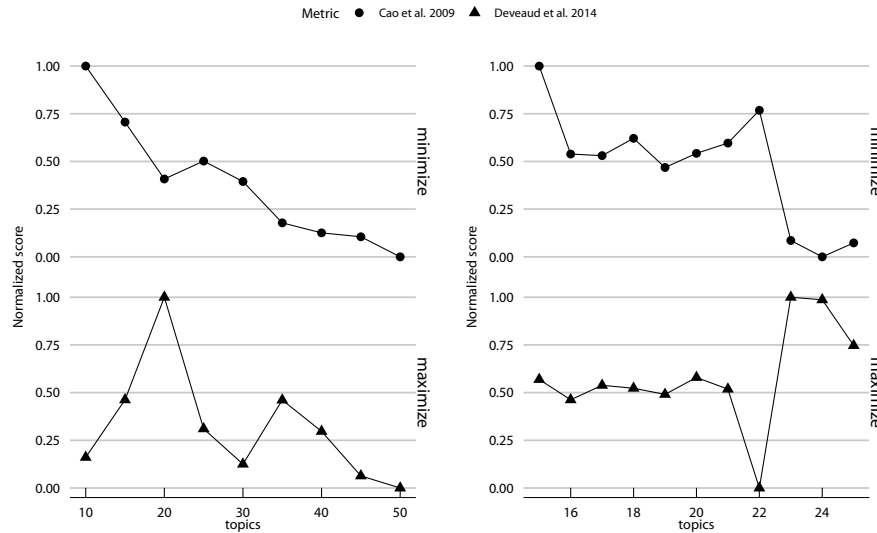


Blue curve shows smoothed LSS values (locally weighted scatterplot smoothing). Shaded ribbon indicates 95% confidence intervals. Grey circles represent LSS scores of individual documents ($N = 2,197$). Dotted vertical lines from left to right: September 28, October 15, and November 26.

The Standard and *The Guardian*. Comparing on the country level, yields equivalent results, where China has a significantly higher mean score than Hong Kong, Taiwan, and UK. The Korean newspapers' relatively small sample size is probably the reason for the results not being statistically significant.

Finally, to test the accuracy of the text model, 36 articles were manually coded. The articles were randomly sampled, but a bias was introduced. For each of the nine newspapers 4 articles were randomly selected. This ensures that each newspaper is included in the sample. The manual coding ranged on a scale from 1 to 5, where 1 corresponds to "violence framing", 3 to balanced framing or absence, and 5 to "peaceful framing" (for coding rules, see Appendix). The agreement between human coding and machine coding was determined by using the Pearson correlation coefficient (Watanabe 2017c, p. 4). The result of $r(34) = .62, p < .001$ is by far not excellent, but still acceptable. The strongest source of disagreement stems from human coded documents with the value 3, where the text model assigned a positive score. A manual inspection of the machine coded articles suggests that the dictionary tends to assign positive scores when terms related to constitutional and civil rights are prevalent in the document (e.g. *China Daily*, August 30, 2014). This means that the interpretability of the text model is not straightforward. While negative and close to 0 values behave as desired, highly positive values include concepts that are not directly related to peaceful behaviors.

Figure 4.3: Topic model optimization metrics.



4.3.3 Cause and Foreign Influence

As described in section 3.3.2, corpus was processed with the lemmatizer and POS-tagger from Stanford's *CoreNLP*, which harmonized the data. Afterwards, common English words (i.e. stop words) were removed using the curated list included in the tools by Welbers, Atteveldt, and Benoit (2017). Then the corpus was transformed into a document-term-matrix with 28,233 unique terms. Initially, relative pruning as suggested by Maier et al. (2018, p. 101) with thresholds of 0.5% and 99% term occurrence, reduced too many terms from the matrix and less than 1,000 were left. Therefore, absolute pruning was applied with a minimum term frequency of 5. This resulted in a matrix with 8,908 terms, which should be sufficiently high for the text model to run reasonably well (Jacobi, Atteveldt, and Welbers 2015, p. 93).

To determine an optimal number of topics for the model, the tool provided by Murzintcev (2020) was used which automatically calculates the model fit metrics suggested by Cao et al. (2009) and Deveaud, SanJuan, and Bellot (2014). The first metric is for minimizing the number of topics and the latter one for maximizing. The optimization procedure was run twice, and the results are plotted in Figure 4.3. First the optimization procedure was run for 10 to 50 topics in 10 step increments, and its result indicated that there is an optimum at 20 topics. Then the process was repeated for 15 to 25 topics in 1 step increments which revealed that the optimal number of topics is estimated at 24.

Then, the keyword assisted topic model developed by Eshima, Imai, and Sasaki (2020) was initialized with 19 undefined topics and the five topics for "foreign influence", "cause", "handover", "rule of law", and "public nuisance" as described in section 3.4.1. The fitted model provides a list of terms that are associated with each topic, and probability values for every document across every topic.

The threshold for the topic probability value of more than 0.5 as recom-

mended by Maier et al. (2018, p. 108) turned out to be too high for the produced model. Only 32 documents would have been classified as containing the cause frame and 0 documents for the foreign influence frame. Therefore, the 90th percentile of all probability values was chosen as a more adequate threshold with 0.143. In order to label the topics and evaluate the model for external validity and semantic quality, the top 20 terms for each topic were inspected. Additionally, 10 random documents per topic were sampled and reviewed. The results of this step are summarized in Table 4.6. The manual inspection process showed that most topics have a solid semantic coherence, especially most of the keyword initialized topics. The semantic scope of the *Rule of Law* topic is much wider than expected. Two of the ten sampled documents just marginally mention terms related to justice or civil rights (*China Daily*, September 30, 2014; *The China Post*, October 1, 2014). For the task at hand, this is not a major obstacle, as this topic was not intended to be used for comparison in the first place. However, it points at improving the keywords for future topic modelling. Another rather unspecific topic is what I labelled “Generation Conflict”. It occurs in a large proportion of documents, but it covers a wide semantic range that is related to terms such as “young”, “old” and generally is associated with documents mentioning opinions of different actors and stories of individuals (e.g. SCMP, June 26, 2014; SCMP, October 2, 2014; *The Guardian*, September 30, 2014; *China Daily*, October 27, 2014). This topic seems to be especially prevalent in the SCMP and *China Daily* with 176 and 55 documents, respectively. Four topics were labelled with “Other” as they had no obvious coherence and can thus be regarded as background topics. The rank-1 score of these topics is also very low (three scoring 0) which is another indicator for them being background topics (Maier et al. 2018, p. 107).

Next, the presence of the cause frame and foreign influence was classified by using the 90th percentile threshold as mentioned above. All documents where the cause topic and foreign influence topic thresholds are higher than 0.143, were classified to contain the corresponding framing. A summary of the results is listed in Table 4.7. The cause frame was found in roughly a third of all documents, where SCMP has the highest absolute count with 219 and *The China Post* the highest relative count with 70.4%. The foreign influence frame was found in less documents with just 161, where also the SCMP has the highest absolute amount with 65.

To demonstrate the performance and semantic coherence of the framing classification, a handful of the randomly selected documents are presented with the keywords highlighted and the corresponding probability value. *The Guardian* reported on the cause before the occupation started:

“The city has been plunged into political crisis after *pro-democracy* activists vowed to take over the streets of the city’s financial district following Beijing’s refusal to grant citizens full *universal suffrage*.”
(*The Guardian*, September 5, 2014, emphasis added)

The above document received a high probability value of 0.62. But also documents with a lower value such as the following with 0.24 have a decent

Table 4.6: Overview of Keyword Assisted Topic Model Results

K	Topic	<i>n</i>	Rank-1	Top-5 Terms
3	<i>Cause</i>	321	0.146	city, government, leader, democracy, election
4	<i>Rule of Law</i>	238	0.108	law, government, rule, people, movement
13	<i>Public Nuisance</i>	49	0.022	road, traffic, causeway, block, area
1	<i>Foreign Influence</i>	37	0.017	media, country, foreign, china, report
2	<i>Handover</i>	29	0.013	british, former, chinese, freedom, 1997
8	Generation Conflict	303	0.138	people, will, one, many, now
5	Police / Violence	263	0.120	police, protester, officer, arrest, teargas
19	Negotiations / Talks	197	0.090	government, student, protest, leung, talk
15	Economic Consequences	195	0.089	market, year, stock, business, financial
12	Umbrella Movement	123	0.056	protest, movement, umbrella, occupy, student
16	US-China Meeting	78	0.036	china, us, president, xi, obama
17	Eviction	73	0.033	court, clear, police, site, injunction
20	Occupy Central	65	0.030	occupy, central, movement, civil, tai
14	Cross-Strait	63	0.029	mainland, taiwan, tourist, protest, visitor
11	Internet Censorship / Cyber Warfare	41	0.019	post, media, online, access, user
7	Pan-Democrats / LegCo Members	30	0.014	lawmaker, legco, member, lai, pan-democrat
10	NPC Decision	28	0.013	committee, basic law, reform, decision, nomination
9	Student Groups / Scholarism	18	0.008	student, class, school, university, boycott
18	UK-China	14	0.006	uk, britain, group, share, museum
6	Anti Occupy	6	0.003	admiralty, protester, occupy, bay, causeway
21	Other 1	26	0.012	rights, worker, government, detain, case
22	Other 2	0	0	hk\$, shop, will, last, new
23	Other 3	0	0	military, pla, china, war, ukraine
24	Other 4	0	0	march, rally, park, victoria, july

K: topic index; *n*: number of documents where topic probability ranked highest; Rank-1: proportion of documents where topic probability ranked highest. Keyword initialized topics are highlighted.

Table 4.7: Summary of topic model based frame classification

Territory	Newspaper	<i>n</i>	Cause Frame		Foreign Influence	
			n	%	n	%
China	China Daily	376	68	18.1	54	14.4
	People's Daily	76	29	38.2	7	9.2
Hong Kong	SCMP	911	219	24.0	65	7.1
	The Standard	381	133	34.9	13	3.4
S. Korea	Chosun Ilbo	17	10	58.8	0	0.0
	Korea Times	21	10	47.6	2	9.5
Taiwan	China Post	152	107	70.4	5	3.3
	Taipei Times	175	109	62.3	7	4.0
UK	The Guardian	88	44	50.0	8	9.1
<i>Total</i>		2197	729	33.2	161	7.3

χ^2 -tests show that the newspapers differ from each other significantly for both frames. Cause frame: $\chi^2 = 254.39, p < .001$. Foreign influence frame: $\chi^2(8) = 44.84, p < .001$. Comparison between countries differs significantly as well: Cause frame: $\chi^2(4) = 225.64, p < .001$. Foreign influence frame: $\chi^2(4) = 35.57, p < .001$.

connection to the framing:

“In a direct challenge to the authority of the nation’s top legislature, the occupiers demanded the withdrawal of the decision by the Standing Committee of the National People’s Congress. This landmark decision was on constitutional reform proposals for the *election* of Hong Kong’s Chief Executive via *universal suffrage* in 2017. The committee has the ultimate authority on this issue.” (*China Daily*, [November 28, 2014](#), emphasis added)

The framing can also originate via a direct quote as this document with a score of 0.66 shows:

“Chow, wearing a black T-shirt with the words ‘Freedom now’ and accompanied by four other student leaders, demanded that the public should have the right to nominate candidates for the 2017 chief executive *election*. ‘The Hong Kong people’s demands for the city’s future constitutional development are very simple civil nominations. We don’t want pre-selected candidates,’ said Alex Chow” (*The China Post*, [October 22, 2014](#), emphasis added)

The document with lowest probability value of 0.14 that passed the threshold also holds decently enough the cause:

“[...] Li Fei has urged pan-democrats to set aside their differences with Beijing on political reform while insisting that the central government will not bow to Occupy Central. In a 50-minute speech

before 800 people from various sectors at AsiaWorld-Expo yesterday, Li called on pan-democrats to ‘let go of prejudice, change their stance and join the “love country, love Hong Kong” camp to actualize universal suffrage.’” (*The Standard*, [September 2, 2014](#), emphasis added)

The foreign influence frame is rarer in the sample and the probability values are overall lower. This document received a value of 0.37:

“‘They are vainglorious in their anti-central government position. They even *conspire* with external forces to undermine national *sovereignty* and ‘one country, two systems,’ he [*Chinese ambassador to Britain Liu Xiaoming*] warned.” (*People’s Daily*, [July 10, 2014](#), emphasis added)

Noticeable, the framing occurred in a quote, which appears also in other documents such as the following one with a score of 0.46:

“The newspaper [*Pakistan Observer*] said the protests are spear-headed by some *foreign* funded elements. Slating the Western media and governments for exaggerating the demonstrations to pressurize Beijing, the editorial said China agreed to govern Hong Kong under the principle of ‘one country, two systems’ but the western media is highlighting the issue unreasonably to advance their sinister designs against China.” (*China Daily*, [October 23, 2014](#), emphasis added)

This pattern of framing via a quote seems to be most common for the foreign influence frame as in this document with a value of 0.19:

“According to the Christian Science Monitor, the Chinese official told his Russian associates, in remarks unpublicized by the Chinese media but reported in Russia, that ‘some Western nations are now supporting the opposition parties of Hong Kong and their goal is to launch a so-called “color revolution” in Hong Kong.’ That is to say, the Hong Kong demonstrations are not just a student-led protest for democracy, but an attempt to overthrow the government, not only in Hong Kong but, ultimately, all of China.” (*The Korea Times*, [October 21, 2014](#))

It also shows that the keywords do not necessarily appear in the document.

4.4 QUOTATIONS

As mentioned previously, the corpus was processed using the natural language processing pipeline *CoreNLP* which also identifies named entities and can detect quotations. The model identified quotes in 1,797 documents which yielded

a rather large and unsorted list of 1,633 speakers. While the model can recognize person names, it is not able to automatically harmonize the various speaker labels and the same person can occur in different ways. For example, former Chief Executive of Hong Kong C.Y. Leung occurred in all possible variations including “C.Y. Leung”, “HK Chief Executive”, “Leung Chun-ying”, and so forth. Therefore, the list of speaker labels was manually summarized by iteratively searching for prominent actors and allocating them to one of the nine actor groups. A conservative attitude was maintained and ambiguous speaker labels such as named students were assigned to the “others” category.

Next, the quotations per actor group were counted on the document level, transforming it into a set of 10 dichotomous variables. An extra variable that indicates whether any quotation is present in the document was generated as well. The abridged result of this step is summarized in Table 4.8 (for the full table, see Appendix). Running several χ^2 -tests indicated that the newspapers significantly differ in quoting the Hong Kong government, protesters, the PRC government and members of the others category. However, the newspapers do not differ significantly in quoting bystanders, as well as the US, UK, or South Korean governments. Consequently, these groups are regarded of lower importance for the comparison below.

The “unknown” category is rather crowded, not only because the language model tends keep falsely allocated hits low, but also because the category includes pronouns that could not be resolved to the original speaker. However, the assumption is that a newspaper article would at least mention the speaker’s name or function once, before it turns to using pronouns. Given the fact that counts for the “unknown” category are very close to the overall counts of quotations but also always lower, hints at such a relationship.

4.5 IN-HOUSE VS BOUGHT COPY

The documents’ sources were determined by using a search term dictionary as described in section 3.2.5. The initial dictionary of press agencies consisted of their full names and their abbreviated forms. Fortunately, the meta information in the raw HTML of *China Daily*, *Chosun Ilbo*, *People’s Daily*, *SCMP*, and *Taipei Times* all contained a designated field for the news articles origin by naming the press agency. This information was appended to the dictionary which resulted in a list of 47 different press agencies. Next, the dictionary search was performed either against the documents’ meta information or main text, depending on the newspaper. In case the document’s meta information did not contain the agency info, the main text was truncated to only include the first and last 100 characters. Only the first hit of the dictionary per document was considered and if there was no hit, the document was considered to be in-house copy. Agencies that occurred in less than 3 documents are gathered in the *Other* category (except *Kyodo* and *Yonhap*). The results are summarized in Table 4.9. Unexpectedly, *Al Jazeera* and TASS did not yield any results.

The validation of the dictionary reports an accuracy of 96.6%, meaning that only 1 document out of 30 was attributed incorrectly. In that case the document

Table 4.8: Summary of quotation counts by actor groups

Territory	Newspaper	<i>n</i>	Quotation		HK Gov		Protesters		PRC Gov		Others		Unknown	
			<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
China	China Daily	376	283	75.27	71	18.88	52	13.83	24	6.38	9	2.39	241	64.10
	People's Daily	76	52	68.42	22	28.95	9	11.84	8	10.53	3	3.95	42	55.26
Hong Kong	SCMP	911	795	87.27	209	22.94	227	24.92	59	6.48	93	10.21	754	82.77
	The Standard	381	293	76.90	54	14.17	63	16.54	18	4.72	14	3.67	259	67.98
S. Korea	Chosun Ilbo	17	15	88.24	4	23.53	7	41.18	0	0.00	2	11.76	13	76.47
	Korea Times	21	19	90.48	2	9.52	1	4.76	3	14.29	2	9.52	17	80.95
Taiwan	China Post	152	104	68.42	21	13.82	38	25.00	13	8.55	13	8.55	88	57.89
	Taipei Times	175	169	96.57	45	25.71	62	35.43	31	17.71	18	10.29	156	89.14
UK	The Guardian	88	67	76.14	22	25.00	27	30.68	10	11.36	9	10.23	65	73.86
	<i>Total</i>	2197	1797	81.79	450	20.48	486	22.12	166	7.56	163	7.42	1635	74.42

Abridged table of quotation counts in documents. Non-significant groups omitted (for full table see Appendix). Several χ^2 -tests were computed for comparison of newspapers. Quotation: $\chi^2(8) = 91.67, p < .001$. HK Gov: $\chi^2(8) = 26.47, p < .001$. Protesters: $\chi^2(8) = 60.42, p < .001$. PRC Gov: $\chi^2(8) = 38.24, p < .001$. Others: $\chi^2(8) = 37.25, p < .001$. Unknown: $\chi^2(8) = 119.60, p < .001$.

Table 4.9: Summary of documents sourced from agencies

Agency	China				Hong Kong				S. Korea				Taiwan				UK		Total	
	China Daily		People's Daily		SCMP		The Standard		Chosun Ilbo		Korea Times		China Post		Taipei Times		The Guardian		n	%
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%				
In-house	309	82.18	8	10.53	872	95.72	223	58.53	0	0	18	85.71	15	9.87	45	25.71	88	100	1578	71.83
Agence France-Presse	0	0	0	0	8	0.88	66	17.32	0	0	0	0	79	51.97	57	32.57	0	0	210	9.56
Associated Press	0	0	0	0	17	1.87	7	1.84	0	0	0	0	38	25.00	15	8.57	0	0	77	3.50
Bloomberg	2	0.53	0	0	2	0.22	0	0	0	0	0	0	0	0	10	5.71	0	0	14	0.64
Central News Agency	0	0	0	0	0	0	0	0	0	0	0	0	15	9.87	0	0	0	0	15	0.68
China Daily	0	0	14	18.42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0.64
Global Times	0	0	17	22.37	0	0	4	1.05	0	0	0	0	0	0	0	0	0	0	21	0.96
Kyodo	0	0	0	0	1	0.11	0	0	0	0	0	0	0	0	0	0	0	0	1	0.05
Reuters	0	0	0	0	10	1.10	2	0.52	0	0	0	0	0	0	34	19.43	0	0	46	2.09
RTHK	0	0	0	0	0	0	72	18.9	0	0	0	0	0	0	0	0	0	0	72	3.28
Voice of America	0	0	0	0	0	0	0	0	17	100	0	0	0	0	0	0	0	0	17	0.77
Xinhua	51	13.56	34	44.74	0	0	3	0.79	0	0	0	0	0	0	0	0	0	0	88	4.01
Yonhap	0	0	0	0	0	0	0	0	0	0	1	4.76	0	0	0	0	0	0	1	0.05
Other	14	3.72	3	3.95	1	0.11	4	1.05	0	0	2	9.52	5	3.29	14	8	0	0	43	1.96

N = 2,197. Column-wise percentages, sums subject to round-off errors. Newspapers' agency sourcing differs from each other significantly: $\chi^2(104) = 5547.81$, $p < .001$. Country differences are significant as well: $\chi^2(52) = 2568.79$, $p < .001$.

was original copy by the newspaper, but it used a photograph by Reuters. The image caption caused the false classification.

The newspapers' difference in agency usage is statistically significant. For most newspapers the majority of published documents is in-house produced. Exceptions are *The China Post*, *Taipei Times*, and *Chosun Ilbo*. *Agence France-Presse* is the most common press agency across the sample. Some press-agencies in the sample are exclusively associated with certain newspapers, such as the Taiwanese *Central News Agency* is associated with *The China Post*. Another notable observation is that the *People's Daily* sources content from *China Daily* and both newspapers also source from *Xinhua*.

Given these insights, it is assumed that for most cases the news agency copy does not have a latent effect on the framing that spans across countries. The newspapers' choices are overall too different to cause latent patterns that could interfere with the analysis. However, there are two exceptions. The less significant, but notable one is the usage of AFP copy by *The Standard*, *The China Post*, and the *Taipei Times*. The other exception is among the Chinese newspapers, as the sampled documents of the *People's Daily* partially relies on news content by *China Daily*.

4.6 PUBLIC ATTITUDES TOWARDS PROTESTS

The Asian Barometer Survey Wave 3 and Wave 4 provide the data source for this aspect of the thesis. Wave 3 includes 19,436 respondents from 13 countries and was conducted in the period from 2010 until 2012. Wave 4 is larger and contains 20,677 respondents from 14 countries, because Myanmar joined the survey. The interviews in Taiwan were conducted in the second half of 2014, in China and South Korea in 2015, and in Hong Kong in early 2016.

First, the datasets were sliced to only include responses from China, Hong Kong, South Korea, and Taiwan, leaving 7,479 observations for Wave 3 and 8,142 observations for Wave 4. Both waves use the same question battery that gauges the respondents' views on the essential characteristics of a democracy. Therefore, no steps were necessary to make both survey responses compatible with each other and thus they can be compared directly. Table 4.10 shows the number of respondents per country who chose the freedom to take part in protests as an essential part of a democracy. On average 11.26% of the selected population in Wave 3 opted for this answer, where the Hong Kongers have the highest proportion with 20.22% and China the lowest with 5.13%. The figures for Wave 4 are very similar with an overall increase across all selected countries, most notably in South Korea. A χ^2 -test was computed to confirm that the differences between the countries are statistically significant.

Next, the "freedom and liberty" index as suggested by Lu (2012, p. 6) was calculated. The four questions were recoded to 0 and 1 values, where 1 means that the chosen answer is associated with freedom and liberty (see section 3.2.6). Then, the sum of the four resulting variables was computed, yielding an index of discrete values ranging from 0 to 4. The mean values and standard deviations for each wave and selected country are presented in Table 4.11. South Korea has

Table 4.10: ABS: Overview of survey respondents choosing freedom to participate in protests as essential

Territory	Wave 3 (2010–2012)			Wave 4 (2014–2016)		
	Respondents	<i>n</i>	%	Respondents	<i>n</i>	%
Hong Kong	1,207	244	20.22	1,217	265	21.77
Korea	1,207	178	14.75	1,200	215	17.92
China	3,473	178	5.13	4,068	228	5.60
Taiwan	1,592	242	15.20	1,657	258	15.57
<i>Overall</i>	7,479	842	11.26	8,142	966	11.86

n: number of survey respondents choosing “people have the freedom to take part in protests and demonstrations” for question item 88 (Wave 3) or 91 (Wave 4). Source: *Asian Barometer Survey* (Wave 3 and Wave 4). Difference between countries is statistically significant for both waves. Wave 3: $\chi^2(3) = 267.16, p < .001$. Wave 4: $\chi^2(3) = 330.54, p < .001$.

Table 4.11: ABS: Overview of Freedom and Liberty Score

Territory	Wave 3		Wave 4	
	Mean	sd	Mean	sd
Hong Kong	0.66	0.87	0.69	0.95
Korea	0.70	0.86	0.84	0.89
China	0.34	0.60	0.34	0.63
Taiwan	0.51	0.84	0.57	0.87
<i>Overall</i>	0.49	0.76	0.51	0.80

N = 15,621. Mean values with standard deviations (sd) of computed score for survey respondents choosing answers matching the concept of “freedom and liberty” (Lu 2012, p. 6). Computed score ranges from 0 to 4 (discrete values). Source: *Asian Barometer Survey* (Wave 3 and Wave 4). Levene’s test indicated unequal variances for Wave 3 ($F = 99.91, p < .001$) and Wave 4 ($F = 130.80, p < .001$). Difference between countries is statistically significant. Wave 3: Welch’s $F(3, 2753) = 99.10, p < .001$. Wave 4: Welch’s $F(3, 2730) = 151.73, p < .001$.

the highest mean scores in Wave 3 and Wave 4 with 0.70 and 0.84, respectively. Respondents from China have the lowest mean index for both waves with 0.34. The differences of the mean values between the countries are statistically significant. Subsequent Games-Howell tests confirm that the pairwise mean values between the countries are statistically significant ($p < .001$), with the exceptions of the pair South Korea–Hong Kong in Wave 3 ($p = .741$).

5 ANALYSIS

The results from the content analysis are compared to each other according to the analytical framework that was developed in section sec. 3. The following section presents each criterion by country and contrasts the corresponding newspapers. As mentioned above, *The Guardian* serves as a reference text, and therefore each criterion is first presented with the results for the British newspaper. For each criterion the newspapers are assessed on whether their coverage is favorable of the Umbrella Movement or not. Quotations of low occurring actor groups such as bystanders are omitted, because they were tested not to be statistically significant. Finally, all results are summarized in a large overview table.

5.1 SALIENCE

5.1.1 *The Guardian*

Among the selected newspapers, *The Guardian* has almost the lowest proportion of documents that cover the protests in Hong Kong with only 0.18% as it is presented in Table 4.1. In absolute terms, the visibility is slightly higher with 88 documents. As Figure 4.1 shows, most of the documents are concentrated in September and October, when the protests were still a novel and the coverage dropped in November with a slight increase in December. However, as Table 4.2 shows, the British daily also has on average the highest wordcounts in the sample. Not only are the articles that mention the protests the longest with almost 800 words, but also in general all articles have a wordcount of 737. These figures match the findings by Sparks (2015, p. 433). Considering this internal difference, it potentially hints at a reporting pattern that does not cover the events very frequently, but in depth. That would align with the findings of Harlow and Johnson (2011, p. 1367), where “Western” newspapers also only occasionally reported about the foreign protests, but with generally longer articles that contained more background information.

5.1.2 *Hong Kong*

Unsurprisingly, the Hong Kong based newspapers have the highest relative visibility. The SCMP mentioned the protests in 3.75% of the documents in the collected sample. *The Standard* has the second highest proportion with 2.37%, but with 381 documents less than half of SCMP in absolute terms. Both newspa-

pers also have the highest coverage leading up to the protests, and thus reported on the public discourse and precursor events.

The Standard typically has shorter articles with an average wordcount of approximately 213, which fits its tabloid-like style. The documents covering the protests are with 279 words about 23% longer in comparison. The difference is more pronounced in the *South China Morning Post* where the typical document length is less than 500 words, but the documents mentioning the protests have on average 766 words which is 55% more. However, the SCMP also has by far the highest standard deviation of 1,011 in this indicator which means that they have very long articles and also very short updates. Considering these differences, the *South China Morning Post* has a much higher visibility than *The Standard*.

5.1.3 China

The *China Daily* has a lower relative coverage with 0.57% than the *People's Daily* with 0.79%. However, in absolute terms the *China Daily* has almost five times more articles mentioning the protests than *People's Daily*. The 376 articles of the *China Daily* are also more evenly distributed across the year and feature a higher coverage in the first half of 2014 than *People's Daily* (see: Figure 4.1). Both newspapers rank in the middle of all outlets regarding this indicator.

China Daily on average has longer articles than *People's Daily* with wordcounts of 397 and 235, respectively. *China Daily's* coverage of the protests is with an average wordcount of 602 about 51% longer compared to its regular document length. The difference of *People's Daily* is more pronounced where the wordcount of the documents mentioning the protests is even 88% higher. While the protests appear to have a higher relative salience in *People's Daily*, the *China Daily* seems to cover the events more regularly and less concentrated.

5.1.4 Taiwan

The two selected Taiwanese newspapers have an overall high coverage of the protests. They do not differ too much in this indicator, where *The China Post* has a slightly higher value with 0.96% than the *Taipei Times* with 0.89%. The two newspapers also have similar figures in absolute terms, *The China Post* has 152 relevant documents and the *Taipei Times* has just 23 relevant documents more. Both newspapers reported on the protests in Hong Kong relatively regularly, with an exception of August 2014 where *The China Post* did not report about it.

Looking at the document length, the wordcount of the *Taipei Times's* regular wordcount of 458 is very close to the overall average of 457. Documents by *The China Post* are on average shorter with a mean wordcount of 348. Documents that mention the Umbrella Movement are not much longer for either of the two outlets. The *Taipei Times* has about 19% longer documents with a wordcount of 546 and *The China Post* has 13% more words with an average of 395.

The salience of the Umbrella Movement is very similar in both newspapers. Where *The China Post* has a higher relative coverage, the *Taipei Times* has longer documents and also more in absolute terms.

5.1.5 South Korea

Among the selected newspapers, *The Korea Times* has the lowest relative number of documents mentioning the protests with only 0.12%. The *Chosun Ilbo* is not much higher with 0.39% and also has with 17 relevant documents the lowest absolute number in this indicator.

The *Chosun Ilbo* also has on average the smallest wordcount with 169 across all selected newspapers. *The Korea Times* in contrast has an average wordcount of 372 which is more than the *Chosun Ilbo* but still comparably low in the sample. Documents which mention the protests are much longer for both newspapers, where *Chosun Ilbo*'s documents have an average wordcount of 309 and *The Korea Times*' documents even have a mean wordcount of 633. This makes the latter one rank as the third highest for this indicator. The difference in wordcount is substantial for both Korean newspapers, where the *Chosun Ilbo*'s documents are 82% longer and *The Korea Times*' are 70% longer.

Neither of the newspapers provide a particularly high salience for the Umbrella Movement. Their reporting frequency and length appears to be similar to *The Guardian*, which is more occasional reporting, but dedicating more space for the issue.

5.2 FRAMING

5.2.1 The Guardian

The news perspective in *The Guardian* appears to be neutral on average with a sentiment score of 0.00 as shown in Table 4.3. This also applies to its balance between peacefulness and violence, where the newspaper ranks with -0.016 exactly in the center of all selected outlets (see Table 4.5). It also has the lowest standard deviation (0.718) for this indicator, which hints at a relatively consistent balance. Table 4.7 shows that half of all relevant document mention the cause of the protesters which also places *The Guardian* sort of in the middle. About 9% of the documents mentioned a potential foreign influence on the protesters which also situates the newspaper more in the center of all selected outlets.

Given these indicators, *The Guardian* seems to have a framing that tends to be neutral. So, it is neither in favor nor against the protesters. Considering the study of Sparks (2015, p. 439), *The Guardian* reported the events with articles that emphasized the acts of violence, but also the general peacefulness of the protesters. Therefore, it appears that the results of these indicators generally agree with previous findings.

5.2.2 Hong Kong

Both Hong Kong based newspapers have a slightly negative news perspective with sentiment scores of -0.01. In terms of violence framing, the SCMP tends to emphasize acts of violence more than *The Standard* with scores of -0.134 and -0.120 respectively. However, both of them are the newspapers with the most

negative scores for this indicator. Meaning that neither is particularly highlighting peaceful aspects of the protests.

With only 24% of all relevant document, the SCMP mentions the cause of the protesters less frequently than *The Standard* with almost 35%. The SCMP is also ranks second lowest among all newspapers for this indicator. However, considering that the SCMP also has more documents in absolute terms, it makes sense that not all articles of the same day mention the cause over and over again. It could be argued that SCMP readers are well aware of the protesters' cause and they only have to be reminded when the cause itself is discussed in the article. A similar point can be raised for *The Standard*, which also mentions the protesters' cause less frequently than *The Guardian*.

The Standard mentioned in 3.4% of all relevant documents a potential foreign influence on the protesters. Also comparing it to the other newspapers, it is the third lowest for this indicator and uses the framing less often than SCMP with 7.1% or *The Guardian*. In Hong Kong, the framing of foreign influences on the protesters also has an additional special meaning. On October 19, 2014, Chief Executive CY Leung publicly stated that he had evidence that "external forces" were involved in the protests (e.g. *People's Daily*, [October 20, 2014](#); *SCMP*, [October 20, 2014](#)). On the following days, this statement was a big topic in the local press during the following days and the framing was also highly contested (Lee and Chan [2018](#), pp. 156-159). With this aspect in mind, it is to be expected that the foreign influence frame would be present in Hong Kong based newspapers.

Neither of the two newspapers exhibits a framing that is clearly in favor of the protesters. Even though, *The Standard* was expected to have a more negative view on the Umbrella Movement, it is actually the SCMP that appears to make use of more critical framing devices. Comparing both papers against *The Guardian*, they appear both to employ a more skeptical view of the protesters.

5.2.3 China

People's Daily has the most negative news perspective among all selected newspapers. *China Daily* score is closer to zero, but still slightly negative with -0.01. Unexpectedly, both newspapers have the most positive scores for the violence framing with 0.449 for *People's Daily* and 0.372 for *China Daily*. Meaning that they both avoid mentioning acts of violence and generally tend to report on peaceful aspects.

The cause frame is not very prevalent in either of the newspapers. *China Daily* has even the lowest percentage of all newspapers with 18.1%, but *People's Daily* proportion is about twice as high with 38.2%. This would speak for both newspapers having a comparably negative framing for this indicator.

China Daily is also the newspaper which most frequently framed the Umbrella Movement as being influenced by external forces with about 14%. The *People's Daily* usage of this framing is very close to *The Guardian* with only 9.2%. In the Chinese speaking press, this particular frame was constantly present among the pro-Beijing newspapers (p. 157). It would also fit the reporting pat-

tern of *China Daily* as it was described by Ha and Shin (2019, p. 211), where the newspaper would blame foreign powers on causing instability in a country.

Considering the values for these indicators, the results are counter intuitive and ambiguous. Besides the violence framing, both newspapers tend to use framing devices that are not in favor of the protesters, especially *China Daily*. However, even though *People's Daily* sources content from *China Daily*, their framing still differs more from each other than anticipated.

5.2.4 Taiwan

The China Post has a considerably negative news perspective with -0.03, which is more at the lower bound of all selected newspapers. *The Taipei Times* on the other hand has with a score of 0.00 the same value as *The Guardian* and is thus more positive than most of the newspapers in the sample. The two Taiwanese newspapers minimally differ in their LSS score. *The Taipei Times* has a mean score of 0.011 and *The China Post* a mean score of -0.082. This small difference is enough that the two newspapers are quite apart when all newspapers' scores are considered. The *Taipei Times* is also in this indicator close to *The Guardian*, but *The China Post* is close to the two Hong Kong based newspapers. Therefore, the *Taipei Times* tends to be more on the balanced side, and *The China Post* tends to emphasize acts of violence.

The cause frame is more prevalent in *The China Post* than in the *Taipei Times* with 70% and 62%, respectively. However, both newspapers most prominently employ the cause frame across all selected newspapers. This suggests that both newspapers have a favorable framing for the protest camp in this indicator.

Both newspapers also strongly refrained from framing the protests as being influenced by foreign powers. Their proportions are almost identical with 3.3% for *The China Post* and 4.0% for the *Taipei Times*. As mentioned above, in Chinese-language newspapers, this framing was highly prevalent. But for the two English-speaking Taiwanese newspapers it appears that this framing device was very uncommon. Thus, this indicator is also favorable of the protesters for both newspapers.

The indicators suggest that *Taipei Times* has a tendency to be favorable of the protest camp. It is less clear for *The China Post*, because it emphasizes violence more and has a less positive perspective. However, the newspaper's usage of the cause frame and foreign influence frame suggest that it still more favorable of the protesters than *The Guardian*.

5.2.5 South Korea

The two South Korea newspapers have a considerable negative news perspective. The *Chosun Ilbo* has a rather negative sentiment score of -0.03 and *The Korea Times* is also negative with -0.02. This suggests that both newspapers are not favorable of the protests for this indicator.

Their framing of violence and peacefulness differs to a great extent. While *The Korea Times* has the third highest average value of 0.289, the *Chosun Ilbo* has with -0.044 a lower score than *The Guardian*. This suggests, that the *Chosun Ilbo*

has a tendency to emphasize violence, but is still rather balanced. But *The Korea Times* can be considered to focus strongly on peaceful aspects of the protests in Hong Kong.

The two newspapers also differ in their usage of the cause and foreign influence frame. While both newspapers rank in the middle for the usage of the cause frame, the *Chosun Ilbo* used the frame in 58.8% of all relevant documents and *The Korea Times* used it only in 47.6%. Thus, the *Chosun Ilbo* employed the cause frame more often than *The Guardian* and *The Korea Times* to a lesser extent. The difference is more pronounced for the foreign influence frame. Among the selected newspapers, the *Chosun Ilbo* is the only one that completely refrained from using it. *The Korea Times* on the other hand mentioned it in 9.5% of all documents which makes it very similar to *The Guardian's* usage. The results for this indicator suggest that the *Chosun Ilbo* has a stronger tendency to have favorable reporting for the protesters than *The Korea Times*.

The difference between the two newspapers is rather clear and pronounced for this criterion. The *Chosun Ilbo* emphasizes violence, but also mentions the cause of the protesters frequently and does not suggest that they are influenced by foreign powers. For *The Korea Times* the statements can be reversed. However, this still does not lead to a clear conclusion which newspaper has for the protesters a more favorable reporting.

5.3 QUOTATIONS

5.3.1 *The Guardian*

As shown in Table 4.8, *The Guardian* has a moderate usage of quotations, where 76% of all relevant documents contained at least one quote. That places the newspaper in the middle compared to all other selected newspapers. In 30.7% of all relevant documents *The Guardian* quoted protesters, making it the most frequently quoted group for this newspaper. The Hong Kong Government was quoted in 25% of the documents and the PRC Government in 11.4%. Looking at the statistically non-significant groups, *The Guardian* quoted bystanders in only one document, and the UK government in two documents. Considering these figures, the newspaper clearly prefers the protesters over the HK government, and provides some space for the PRC government.

5.3.2 *Hong Kong*

The Standard used quotations in 76.9% of all relevant documents, the SCMP on the other hand uses quotes more often in about 87% of the relevant documents. Both newspapers have the same ranking for the most important groups. The protesters are the most frequently quoted group. However, *The Standard* has overall lower proportions for quotations than SCMP. While the *South China Morning Post* included a quote by protesters in almost 25% of all relevant documents, *The Standard* included one in only 16.5%. The second most commonly quoted group is the HK government which the SCMP quoted more frequently than *The Standard* with 22.9% and 14.2% respectively. The third group is the

PRC government that was quoted in 6.5% of all relevant documents by SCMP and in 4.7% of the relevant documents by *The Standard*. Considering other groups, the SCMP has the highest diversity of quotes in the selected newspapers. It quoted bystanders in 3.25% of all relevant documents, counter protesters in 2.85%, the UK government in 2.3%, and other governments (e.g. US) in 3.62%. *The Standard* shows a similar variety, but with overall lower proportions.

While both newspapers have similar quoting priorities, *The Standard* uses less quotations than SCMP. Considering that *The Standard* overall usage of quotations is very similar to *The Guardian*, its preference to quote protesters or the HK government is much lower. For this criterion both newspapers are considered to be neutral towards the protesters. The difference in percentages between quotes by the protesters and the government are very small.

5.3.3 China

The two mainland Chinese newspapers have a lower usage of quotes than *The Guardian*. *People's Daily* only included a quote in 68.4% of all relevant documents which makes it the lowest ranking newspaper together with the Taiwanese *The China Post*. *China Daily* includes quotes about as often as *The Guardian*. The China-based newspapers have the same prioritization for the largest groups, where they most frequently quote the HK government. While *China Daily* only quoted the HK government in 18.9% of the relevant documents, the *People's Daily* quoted it in 29% of all relevant documents which makes it also the newspaper with the strongest preference for this group. The protesters were quoted in 11.8% of the *People's Daily* documents and the PRC government in 10.5%. *China Daily* has a similar weighting for the protesters with 13.8% but quotes the PRC government less frequently with 6.4%. Such patterns of *China Daily* were also found by Ha and Shin (2019, p. 209).

Both newspapers clearly show a preference for the HK government in this criterion. Especially, *People's Daily* has a much stronger bias than *China Daily* for the indicators.

5.3.4 Taiwan

The *Taipei Times* has the highest usage of quotations among all selected newspapers and as mentioned above, *The China Post* the lowest. However, their prioritization is similar, but also comes with different weights. Both newspapers quote protesters most frequently where the *Taipei Times* quotes this group in 35% of all relevant documents and *The China Post* quotes them in 25%. The HK government is quoted in 25.7% of the *Taipei Times*' documents and in 13.8% of *The China Post*'s documents. The *Taipei Times* has the highest proportion of documents that quote the PRC government with 17.7%. *The China Post* also has a relatively high usage of PRC government quotes with 8.6%, but less than *The Guardian*. It should be noted here, that both newspapers had occasional quotes of the Taiwanese KMT government and also the DPP opposition parties. However, both groups were found in less than 4 documents for each newspaper.

Considering the quoting prioritization, both newspapers have favorable coverage of the protesters for this criterion. However, it is unexpected that the *Taipei Times* has the highest usage of quotes by the PRC government.

5.3.5 South Korea

Both South Korean newspapers have an overall high usage of quotes. *The Korea Times* uses quotes in 90.5% of all relevant documents and the *Chosun Ilbo* just a bit less in 87.3%. The two newspapers differ strongly in their quoting preferences. The *Chosun Ilbo* quotes the protesters in 41.2% of all relevant documents which also makes it the newspaper with the strongest usage of this group's quotes. It quotes the Hong Kong government in 23.5% of all relevant documents, and the PRC government is not quoted at all. The order is reversed in *The Korea Times*, where the PRC government is quoted in 14.3%, the HK government in 9.5%, the protesters in only one document (4.76%) of all relevant documents. The latter group is also least frequently quoted by *The Korea Times*.

This strong difference between the two newspapers leaves the conclusion for this criterion that the *Chosun Ilbo* is strongly in favor of the protesters and *The Korea Times* strongly against the protesters.

5.4 IN-HOUSE VS BOUGHT COPY

5.4.1 The Guardian

As mentioned in section 3.2.5, this indicator is just to control for potential shared patterns by two different newspapers. As the results in section 4.5 suggest, most of the newspapers use different sources with few important exceptions that are highlighted here. Most notably the Chinese newspapers' usage of *Xinhua* copy and also the Taiwanese newspapers' dependence on Western agencies such as AFP and AP.

All collected documents by *The Guardian* are in-house copy and is not attributed to any external agency. Considering that the documents were retrieved via their official developer API, it is reasonable to assume that the outlet can only offer documents for download for which they have the copyright. A manual inspection of *The Guardian* website also shows that all texts are indeed produced by the outlet, but the images are occasionally attributed to external agencies. While this is still speculative, it can be concluded that the frames found that the analyzed documents published by *The Guardian* all reflect the news outlet's viewpoint. But it is possible that *The Guardian* also relied on external sources at the time of reporting, but just does not include that in their online archives.

5.4.2 Hong Kong

The *South China Morning Post* sources only 4.3% of the relevant documents from press agencies. *The Standard* on the other hand sources about 41% of their content from external outlets. The largest share of external content is by the

public radio station RTHK with 18.9%, closely followed by *Agence France-Presse* with 17.3%.

5.4.3 China

China Daily produces most of its content in-house. Only 17.8% of the relevant documents were attributed to an external organization. The *People's Daily* on the other hand sources almost 90% of its content externally. As mentioned before, both Chinese newspapers use copy from the state agency *Xinhua*. But while around 13% of the relevant documents by *China Daily* were attributed to *Xinhua*, *People's Daily* uses the copy by the state-agency in almost 45% of the relevant documents.

5.4.4 Taiwan

The *Taipei Times* and *The China Post* both quite strongly rely on agency copy with 74.3% and 90.1%, respectively. Both use AFP as their main source, and the relevant documents by *The China Post* are even to almost 52% attributed to the French agency. The *Taipei Times* attributes almost one third of the relevant documents to AFP. Both newspapers also source content from *Associated Press*, where *The China Post* relies more strongly on the agency with 25% and the *Taipei Times* less with only 8.6%. Another notable pattern is that *The China Post* sources content from the Taiwanese state-sponsored *Central News Agency*, but the *Taipei Times* does not. The *Taipei Times* instead sources from Reuters which in turn is not a used agency by the other newspaper.

5.4.5 South Korea

The two South Korean newspapers differ rather strongly in their content sourcing. While *The Korea Times* produces most content in-house and only sources 14.3% from external agencies, all relevant documents by *Chosun Ilbo* are attributed to the broadcaster *Voice of America*. The radio station is funded by the U.S. Congress and distributes its content in more than 40 languages (Voice of America Office of Public Relations, [n.d.](#); Wang and Hong 2011, p. 344).

5.5 PUBLIC ATTITUDES

5.5.1 Asian Barometer Survey

The final criterion for comparison is an external point of reference. The *Asian Barometer Survey* serves to gauge the newspapers' alignment with public opinion on protest participation in their respective countries. Specifically, the survey respondents' understanding of a democracy's essential characteristics as introduced by Lu (2012, p. 6) are considered. The results are presented in Table 4.10 and Table 4.11.

5.5.2 *Hong Kong*

The respondents from Hong Kong have overall the most favorable attitude towards protests among the selected countries. Over 20% of all respondents chose “freedom to participate in protests” as an essential characteristic. This share increased from Wave 3 to Wave 4 to almost 22%. In terms of their understanding of “freedom and liberty” as an essential characteristic, the mean is the second highest with 0.66 in Wave 3 and 0.69 in Wave 4. These observations are in line with previous studies which suggest that Hong Kong citizens value the right to protest (Lee and Chan 2018, p. 26-27).

5.5.3 *China*

Mainland Chinese respondents have the least favorable attitude towards the freedom to protest. Among the respondents, only about 5% chose the freedom to protest as an essential characteristic, making it the group with the lowest share in this point. Similarly, with 0.34 they are also the group with the lowest mean for the “freedom and liberty” index, that has also not changed between the waves.

5.5.4 *Taiwan*

Taiwanese respondents have a moderate view on the freedom to protest, where about 15% chose the corresponding survey item. In terms of their views on “freedom and liberty”, the Taiwanese have the second lowest mean with 0.51 in Wave 3 and 0.57 in Wave 4. This concludes that the Taiwanese are generally have a favorable attitude of protests.

5.5.5 *South Korea*

The South Koreans have a similar view on the freedom to protest as the Taiwanese. In Wave 3, also almost 15% chose the freedom to protest as an essential characteristic. However, unlike the Taiwanese, the increase to Wave 4 is stronger, where almost 18% chose the option in the survey. A similar increase is also observed in their mean “freedom and liberty” index. While in Wave 3 it was 0.7, in Wave 4 it increased to 0.84. Based on these figures, the South Koreans have a very favorable view on the right to protest (J.-E. Lee 2012, p. 412).

5.6 SUMMARY OF ANALYSIS

Table 5.1 gathers all results that were discussed above. Each criterion was ranked on a five-point scale ranging from strongly in favor (+ +) over neutral (o), to strongly against (– –).

Table 5.1: Summary of analysis.

	China		Hong Kong		South Korea		Taiwan		UK
	China Daily	People's Daily	SCMP	The Standard	Chosun Ilbo	Korea Times	China Post	Taipei Times	The Guardian
Salience									
Coverage	0.57%	0.79%	3.75%	2.37%	0.39%	0.12%	0.96%	0.89%	0.18%
Document Length	397.3	235.9	493.9	213.5	169.6	372.6	348.6	458.4	737.3
Relevant Document Length	602.4	442.4	766.1	279.1	309.1	633.5	395.0	546.8	798.5
<i>Result</i>	0	+	++	+	–	--	+	+	–
Framing									
News perspective	-0.01	-0.04	-0.01	-0.01	-0.03	-0.02	-0.03	0.00	0.00
Violent vs peaceful	0.372	0.449	-0.134	-0.120	-0.044	0.289	-0.082	0.011	-0.016
Cause	18.1%	38.2%	24.0%	34.9%	58.8%	47.6%	70.4%	62.3%	50.0%
Foreign Influence	14.4%	9.2%	7.1%	3.4%	0%	9.5%	3.3%	4.00%	9.1%
<i>Result</i>	0	0	–	–	+	–	+	+	0
Quotations									
HK Government	18.9%	29%	22.9%	14.2%	23.5%	9.5%	13.8%	25.7%	25%
Protesters	13.8%	11.8%	24.9%	16.5%	41.2%	4.8%	25%	35.4%	30.7%
PRC Government	6.4%	10.5%	6.5%	4.7%	0%	14.3%	8.6%	17.7%	11.4%
<i>Result</i>	–	--	0	0	++	--	+	+	+
Sources									
Bought copy	17.8%	89.5%	4.3%	4.7%	100%	14.3%	90.1%	74.3%	0%
Public Attitudes									
Freedom to protest		5%		21%		16%		15%	N/A
Freedom and Liberty		0.34		0.68		0.77		0.54	N/A
		--		++		+		+	

++: strongly in favour of the protesters; +: in favour; 0: neutral / ambiguous; –: against;
 --: strongly against.

6 DISCUSSION

The results of the analysis are discussed in this section. First, the weaknesses of the analytical framework and content analysis techniques are highlighted. Next, the newspapers are contrasted on the country level, where each country is discussed separately. Then, the similarities and differences between the newspapers are discussed across countries by using the concept of the protest paradigm. Finally, the patterns found in the different countries are compared for similarities and differences. Throughout these discussions, qualitative samples are presented from the scraped data to illustrate the patterns.

6.1 LIMITATIONS

The comparison of reporting patterns in English-language outlets does of course not cover the complete variety of mediated discussions on foreign (or domestic) protests. The views on the Umbrella Movement in each country are more complex than the results of this analysis suggest. Also using *The Guardian* as a reference text worked out in this case, but other newspapers might have been a reasonable selection as well. Additionally, this thesis is only concerned with the textual components of the news coverage, but not the imagery used. The selection of pictures can create another level of complexity to the framing of protests and enhances the written information greatly.

Furthermore, the performance of some text models is comparatively weak. The sentiment analysis as proxy for news perspective has the major weakness that the results were not statistically significant. Probably, the text model is too generic for the task it was used here. Protests are rich in conflict and debate, and therefore, the words used for reporting on such events is generally negative and the subtle nuances could not be captured by the sentiment dictionary. The moderate performance of *Latent Semantic Scaling* also produced partially counter intuitive results. While the recognition of terms relating to violence worked very well, the opposite end of the text model is too fuzzy and included terms that are not directly related to peaceful behaviors. As the manual inspection and validation coding showed, some documents scored positively that included terms relating to lawful behavior, constitutional rights, and democracy. Consequently, documents by *China Daily* were scored with positive values where they should be closer to 0, as the emphasis on rule of law is rather prominent in said documents. Additionally, it might be possible that the training corpus skewed the results of the text model. It mainly consists of Western online news, and thus might include a majority of documents that in-

herently connect the terms “pro-democracy”, “civil rights”, and “peaceful” with each other. The negative score of the term “police” is a potential indicator for such a bias in the corpus. However, systematic experiments would be required to investigate the behavior of the model. Furthermore, the text model is not fine-grained enough to determine the stated reason for violence. Acts of violence by protesters or by police officers are treated the same. Of course, all the numbers generated by the text models have imperfections and are not expected to be accurate. However, it is assumed that the overall direction is correct and that wrong classifications on the document level are balanced out with a higher level of aggregation.

The Keyword Assisted LDA model performed quite well within its own limitations. While it can identify that certain frames are present in a document, it cannot determine the associated valence and context. Even though a document mentions the cause of the protesters, or suggests the influence of “external forces”, it does not tell us how it is presented.

In order to enhance the numbers generated by the text models, the discussion below includes selected excerpts from documents that illustrate certain points well. It also helps to connect the rather abstract numbers with the patterns that are being investigated.

Another issue is posed by the unequal sample sizes per newspaper. For example, SCMP and *China Daily* both had hundreds of documents, *Chosun Ilbo* and *The Korea Times* each had less than 25. This makes it very difficult to draw clear conclusions for the South Korean newspapers.

Finally, the survey batteries of the ABS are also not ideal. It only provides a very small glimpse into public opinion, and it should not be interpreted for more than it is.

6.2 COUNTRY-WISE DISCUSSIONS OF NEWSPAPERS

The Guardian appears to have taken a relatively neutral perspective on the protests in Hong Kong. The coverage is mostly driven by events as shown in Figure 4.1. But there was also reporting on the developments before the actual protests started. This conforms with the findings by Sparks (2015, p. 431) who highlights that *The Guardian* was among the newspapers with comparatively high coverage. It is rather unexpected though that *The Guardian* would relatively often employ a foreign influence framing. According to Sparks (2015, p. 435) UK newspapers were debating about the UK's role in the protests and their responsibility towards the Joint Declaration that was made before 1997. Maybe this aspect of the debate leaked into the foreign influence framing, where the protesters were framed to be motivated by such agreements. However, Sparks (2015, p. 437) also notes that *The Guardian* used favorable language towards the protesters in the sample he used. This aspect cannot be confirmed with the findings of this study, but also not be rejected.

In terms of the criteria the two Hong Kong based newspapers do not differ too much. The strongest difference is in their overall coverage of the protests, where the SCMP intensively reported on the Umbrella Movement and *The Stan-*

dard measurably less. The reporting by SCMP in particular cannot be described as biased towards one group. It featured a large variety of quoted actors, including foreign governments and counter-protest groups. Probably, the news value of international opinions on protests in Hong Kong are considered to be very high by the SCMP editors. The newspaper might be described as an aggregator on viewpoints. The editors probably also considered that their readership mainly consists of foreign nationals working in Hong Kong and thus aim to make the news relevant to them as well. It also appears that SCMP aimed at providing a balanced reporting on the events. Lee and Chan (2018, p. 196-195) note that the journalists in Hong Kong try to maintain a high-degree of professionalism, but also that the mainstream media “exhibited an overall bias against the movement”. This would fit the observations made here reasonably well.

For both newspapers, the cause of the protesters was not very prevalent. Probably, because the readership is expected to know the context of protest camps that they physically encounter on their way to work. Public nuisance however was an important frame in the domestic coverage of the protests (p. 153). While it was not measured in this analysis, it might be possible that the difference between the two newspapers could be stronger in this aspect.

While the Hong Kongers’ diffuse support for the freedom to protest is relatively high, detailed surveys suggest that the specific support of the Umbrella Movement was not particularly high. Before the protesters occupied central, the public approval was quite low at around 25 to 30% (p. 150). Only after the protests began the support increased to 35%, but stagnated due the course of the Umbrella Movement (p. 150). People not supporting the movement remained stable and robust (p. 150). Considering these factors, it appears that the two newspapers align with public opinion reasonably well.

The two Chinese newspapers do not differ too much, neither has a favorable view on the Umbrella Movement. The *People’s Daily* appears to be stronger in their views than *China Daily*. But while the former has an event-driven reporting pattern, the latter reported consistently over the year. As mentioned previously, the results of the LSS text model are counter intuitive. Nevertheless, the strong avoidance of both newspapers to focus on a violence framing of the Umbrella Movement can also be explained by considering the target audience. The newspapers want to create a favorable image of China and stating how one of their city experiences violent clashes would go against that idea. It generally does not produce a good image to stress how the local law enforcement attacks young people protesting for more political rights. *China Daily* appears to have used a different strategy for framing the events by frequently quoting official sources. This would align with the expected patterns of the protest paradigm. Additionally, as the manual inspection of documents hints, the Chinese newspapers stressed the importance of rule of law, and frequently called the protests “illegal”:

“By assisting court officials to clear streets blocked by the illegal campaign and arresting those who obstructed them, the police performed its statutory duties and upheld the rule of law.” (*China Daily*, November 28, 2014)

Considering the results of the ABS, it appears that the two Chinese newspapers and public opinion align with each quite well. Not only are the newspapers skeptical about the Umbrella Movement, also the survey respondents would not set the freedom to participate in protests as a priority.

The Taiwanese newspapers appear to be very similar in their reporting. They frequently remind the readers of the protesters' cause and they also have the same priorities in their quotations. Further, they tend not to emphasize violent behaviors too strongly. Given that the two newspapers allegedly have opposing political orientations, the differences are less strong than expected. A subtle difference was found in the *Taipei Times* which also quoted the Taiwanese opposition party DPP. Considering that the Taiwanese newspapers also both strongly rely on external agencies, it also appears that their views align more strongly with "Western" media. Both main sources were AFP, AP, Reuters, and Bloomberg, which all can be considered as "Western" agencies. The subtle difference is that the *The China Post* also used the Taiwanese *Central News Agency*.

While the Taiwanese respondents of the ABS are not as enthusiastic about the freedom to protest as the Hong Kongers, it can be said that they value it. This would also align with the newspapers' generally favorable coverage of the protests in Hong Kong.

The two South Korean newspapers exhibit the strongest contrast in their reporting patterns. The *Chosun Ilbo* seems to be overall in favor of the protesters. It favors them in terms of quotations and does not quote the PRC government at all. The readers are frequently reminded of the cause and the foreign influence framing was also completely missing in the sample. *The Korea Times* on the other hand is overall against the protesters which is reflected in the usage of the foreign influence frame and also the very small share of quotes by protesters. It also mainly used original copy for covering the protests, which is also the opposite of *Chosun Ilbo*. However, both newspapers are strongly event driven and only occasionally reported on the protests at all. One effect of this shared reporting pattern is that the sample size is almost too small to draw definite conclusions for either of them. Considering that the overall number of relevant documents for the two South Korean newspapers is very small, it appears as if the newspapers did not want to provide too much space for the issue. Probably, the two newspapers would prefer to draw their readers' attention to issues directly relating to Korea.

The South Korean views on the freedom to protests are very similar to the Taiwanese respondents. Considering the newspapers, it appears that the *Chosun Ilbo* tends to be more aligned with public opinion on this matter. However, the analyzed content by *Chosun Ilbo* is entirely sourced from *Voice of America* which is expected to show Western reporting patterns. Additionally, both newspapers allegedly have a conservative political orientation. This could hint that there were no firm opinions on this issue within the conservative camp.

6.3 PROTEST PARADIGM

Revisiting the protest paradigm, most of the conclusions align with results suggested by previous studies. Regarding quotation preferences, the two mainland Chinese newspapers and *The Korea Times* clearly have a preference for official sources over protesters. This is a clear indication that the three conform with the paradigm in this aspect. These three outlets also tended to frame the protests unfavorably by rarely stating the protesters' cause and hinting the influence of foreign forces on the protesters. This kind of framing undermines the legitimacy of the protesters, because it makes their intentions appear non-genuine. However, the three newspapers' depictions of violence do not seem to match with the expectations. All of them received very positive scores by the LSS text model. Besides the limitations that were highlighted above, it also is possible to explain this pattern by considering the potential images created by acts of violence. Such acts could shed a negative light on the law enforcement and authorities. Reporting on how protesters behave violently, and police officers react to it, might create an image where the government lost control over the situation. Keeping in mind that the protesters' occupation lasted for more than two months, continuous reporting on violence by protesters would imply that the government is weak and tolerates such behaviors. It appears that the three newspapers used a different framing strategy instead, where they often emphasized the importance of the law, and that the protesters should respect it. For example, an excerpt from a document that received a score of 0.70 contains this statement:

“Commenting on the ongoing protests in the SAR, Hua Chunying said the central government strongly opposed illegal activities which harmed the rule of law and social stability in Hong Kong. Hua said she was confident the Hong Kong SAR Government would handle the issue in accordance with the law and maintain social stability in the city.” (*China Daily*, [August 30, 2014](#))

This illustrates how the framing strategy includes reiterating the importance of the law as the basis for decision making in Hong Kong. Interestingly, similar framing appears in this *The Korea Times* document that received an LSS score of 1.92:

“Democracy is primarily about its institutions, and only secondarily about suffrage and candidates. [...] As pro-democracy protests continue in Hong Kong, they show the tendency to focus on the procedural aspects of democracy. Undisputedly, they are important. However, the more substantive questions relate to freedom and the rights of the people, individuals and minorities. From an outside perspective, these questions are being left out of the discourse. Pro-democracy protesters might be right about their procedural claims; however, the Hong Kong authorities do seem to have a point regarding individual rights.” (*The Korea Times*, [October 21, 2014](#))

The two Hong Kong based newspapers partially conform with the protest paradigm, but not to an extent as the previously discussed outlets. SCMP's and *The Standard's* quotation preferences provide a reasonable amount of space for various parties, but vaguely tend to favor the protesters in this regard. As discussed above, the cause frame is not frequently present, and it is assumed that this is owed to the fact that the readers are aware of the events. The foreign influence frame is found in both newspapers, but with the notion that there was a larger debate around this issue at in October 2014. However, both newspapers have the strongest emphasis on violence where both of them have the most negative mean LSS scores. This would fit the reporting pattern of emphasizing disruption and exaggerating violence very well (Lee and Chan 2018, p. 31). Both newspapers are very similar in terms of the protest paradigm, even though they allegedly have different political orientations. The SCMP is generally more critical of China and *The Standard* more friendly towards Chinese integration (Chan 2002, p. 102-p.103). It appears that their convergence stems from similar reasons. Their readers are mostly Western foreigners working in Hong Kong (*The New York Times*, March 31, 2018; Chan 2002, p. 102). The Umbrella Movement caused a major disruption in the occupied area where mostly banking professionals were affected, but also small business owners (Lee and Chan 2018, p. 233). While the SCMP readers are more likely to belong to the former group, *The Standard* aligns more with the latter. As noted, *The Standard* focuses more on "small" stories:

"Meanwhile, in Causeway Bay, a staffer in a shop selling bird's nest and Chinese herbs said their business had dropped by around 80 to 90 percent. But despite the grumblings from shop owners, the protesters seem keen to carry on with their struggle. [...] A tourist from Sichuan, surnamed Li, said she was afraid of the protest even though she was told it was a peaceful rally in Tsim Sha Tsui. Another shopper from Shenzhen, surnamed Huang, said the movement has affected her group's shopping. But protesters in Causeway Bay remained upbeat, while admitting it would be hard to sustain the occupation over a long period." (*The Standard*, October 3, 2014)

This story also frames the protest in terms of economic consequences that Lee and Chan (2018, p. 161) and Feng (2017, p. 563) included in the "public nuisance frame". The SCMP also used this framing, but appears to focus more on the bigger picture and concluding that the protest did not affect the economy too strongly after all (SCMP, December 4, 2014)

The more recent findings on the protest paradigm suggest that the ideological congruence between the protest group and the news outlet influence the framing (Kim and Shahin 2020, p. 402). The *Chosun Ilbo*, the two Taiwanese newspapers, and *The Guardian* had a preference for quoting the protesters. With the exception of *The China Post* they quoted the protesters in more than 30% of the relevant documents and provided less space for other groups. These are also the newspapers that frequently used the cause framing and did not use

the foreign influence frame often. Additionally, just because the foreign influence frame was classified by the LDA, it does not mean that the frame remains uncontested:

“Last Friday, the overseas edition of the official People’s Daily newspaper published a front-page article accusing the National Endowment for Democracy, a nonprofit organization funded by the U.S. government, of involvement in the demonstrations in Hong Kong [...] The State Department has rejected point-blank charges that the U.S. government was ‘manipulating the activities in any way of any person, any group, or any political party’ in Hong Kong.” (*The China Post*, [October 15, 2014](#))

While these pieces fit together nicely, the depiction of violence is an inconclusive affair. All three outlets’ mean scores are close to 0, where the *Chosun Ilbo*, *The China Post*, and *The Guardian* have negative mean scores. The *Taipei Times* has a slightly positive score. A potential explanation is that these newspapers tend to report the aspects of violence as much as they report on debates and the daily activities at the protest camps.

The framing strategy could be to suggest that all events are presented objectively without showing a particular preference for any side. The preference for the protesters is then created by allowing them to explain the reasons for their behaviors. The readers are then to decide whether the cause justifies the means. Another explanation is that the newspapers report on violence in order to invoke sympathy for the protesters. One article illustrates this point that scored -0.82:

6.4 DIFFERENCES AND SIMILARITIES BETWEEN COUNTRIES

Looking at the bigger picture, some observations can be made on how the different countries’ media systems covered the Umbrella Movement. Considering Figure 4.1, the relative visibility of the events decreases starting from Hong Kong with the highest visibility, over Taiwan and China, to South Korea and UK with the lowest relative coverage. It appears that there are at least two factors that influence the salience of the protest. The first one being the geographical distance. China and Taiwan are both geographically very close to Hong Kong and both have a decent coverage of the protests. South Korea is further away, and the coverage decreased more. Finally, the UK is most distantly located from Hong Kong and had the overall lowest relative visibility. However, for the UK only one newspaper was selected.

The second factor is blurry and speculative as there are many possible explanations but could be summarized under “cultural” or “political” proximity. The Umbrella Movement had a higher visibility in the Taiwanese newspapers than in the mainland Chinese ones, even though it is geographically further away and has no jurisdiction over the Special Administrative Region. So, it might be possible that political factors come into play that are inward and outward oriented. On the one hand, the Umbrella Movement was of special interest for

Taiwanese voters in the context of the Sunflower Movement (Ho 2019, p. 5). Both were student movements that challenged a government decision and used civil disobedience as their means (p. 2). Such framing can be found in the *Taipei Times*:

“In the view of escalating violent threats and attacks on peaceful demonstrators in Hong Kong, Taiwanese voters should assert their democratic right of self-determination and reassess rapprochement with China.” (*Taipei Times*, October 8, 2014)

On the other hand, the English-language newspapers are probably not too concerned in addressing domestic readers and to influence voters but focus more on foreign audiences. The Umbrella Movement was protesting a decision by the PRC’s National People’s Congress and therefore, against the mainland government (Ho 2019, p. 60; Lee and Chan 2018, p. 3). From a political standpoint, the Umbrella Movement and Taiwan have a common enemy in Beijing. Thus, it appears to be important for the Taiwanese newspapers to create international attention for protests against the Communist party rule in China. This might also explain the lack of pluralism in the coverage by the selected Taiwanese newspapers. It appears that the two newspapers somewhat united in the agenda to discredit CCP rule:

“Finally, with democracy increasingly becoming a common pursuit among Asian nations, the Chinese government is advised to trust in the wisdom of its citizens, and reassess the need for the required reforms, as democracy remains the ideal foundation towards prosperity.” (*The China Post*, October 1, 2014)

Similarly, there are statements that contextualize the protests with China’s foreign affairs:

“The political tumult in Hong Kong has become a headache for Beijing at a time when party leaders are grappling with a host of challenges, from a slowing economy to diplomatic skirmishes with neighbors like Japan, Vietnam and the Philippines.” (*Taipei Times*, October 7, 2014)

A factor that leads to a lower visibility in mainland Chinese newspapers, could be the deliberate attempt not to give too much attention to the protests. As mentioned above, it is reasonable to assume that the English-language newspapers from China prefer to create a positive image of their country (Rawnsley 2015a, p. 465). Also, the strong alignment of the *People’s Daily* and *China Daily* is probably best explained by the tight level of control by the PRC government on the two outlets and the media system as a whole (Zhou 2015, p. 4). It also reasonable to assume that the PRC government prefers to disseminate their statements and viewpoints in the international media sphere, and therefore, they are together with the HK government the most quoted group in the Chinese

newspapers. This matches the study by Ha and Shin (2019, p. 211), where mainland Chinese journalist also exhibited a preference for quoting official sources, especially Chinese authorities.

The newspapers in Hong Kong, Taiwan, UK and South Korea (except *The Korea Times*) showed a preference for quoting the protesters. Considering that news media framing also includes the function of problem definition (Entman 1993, p. 52), quotes can be used by journalist to enhance this aspect of framing (Brüggemann and D'Angelo 2018, p. 99). Using the protesters own words to define the problem and explaining their cause does not only establish more authenticity, but also gives them more legitimacy (Boyle and McLeod 2018, p. 305). While journalists might not necessarily use quotes by protesters in their favor, they use them to show their readers that they are where “the action is”. These practices can be tracked back to the set of values that the journalists subscribe to (Ryfe 2019, p. 295). Quoting protesters is part of the professional value to present different viewpoints that to show the readers that the information is first-hand and authentic. As it appears, this is a contrast to the mainland Chinese news outlets where the problem definition is handed to government authorities.

With the data of this study, it can be said the English news media in South Korea exhibit a great amount of diverse views in their coverage of the Umbrella Movement. The two selected newspapers differed to a large extent in their framing, even though they are ascribed to have a similar political orientation (J.-w. Lee 2011, p. 1; Ha and Shin 2016, p. 541). However, the visibility of the protests was also very low. This low salience points at a low relevance of the Hong Kong protests for domestic affairs in South Korea. The low relevance in combination with the issue being relatively new, might have left it open for interpretation and contextualization by the news outlets which enabled it to be discussed with very different viewpoints. As Ha (2015, p. 755-756) notes, South Korean newspapers presented the Arab Spring in the context of South Korea's national interest and concern. The *Chosun Ilbo* “domesticated” the Arab Spring by contextualizing it with North Korea (Ha and Shin 2016, p. 550). Such a pattern cannot be observed with the English edition of the newspaper in the case of the Umbrella Movement. This is most likely owed to the fact that all relevant documents were sourced from *Voice of America*. However, *The Korea Times* exhibits commentary that contextualizes the protests with Korean history:

“While China followed through on some of the pragmatic economic reforms initiated by Deng Xiaoping, there would be no democracy. Korea introduced democratic elections, but Chinese leaders ruled by intimidation and force.” (*The Korea Times*, October 2, 2014)

The framing also goes into the direction to discredit communist governments:

“The conservative government of Korea has no desire to offend Beijing by speaking out in defense of demands for democratic

elections even though Korean democracy rose from mass protests against authoritarian rule. And Korean far-leftists amazingly seem on occasion to sympathize with the dictatorship of North Korea, where the least sign of protest means torture, imprisonment and death. [...] Although the central government, ruled mysteriously, secretly, under the thumb of the Communist Party, might appear all-powerful, the country is always in danger of fragmentation despite rising economic prosperity and increasing military displays from the Yellow Sea to the East China Sea to the South China Sea.” (*The Korea Times*, [October 23, 2014](#))

While the South Korean news domesticated the protests in Hong Kong to a small extent, this phenomenon was more prevalent in the Taiwanese newspapers. As mentioned before, it appears that in Taiwan the protests were mainly contextualized with the *Sunflower Movement*. Especially the *Taipei Times* frequently¹ established the connection between the two movements:

“There are similarities between Hong Kong’s Occupy Central and Taiwan’s Sunflower movement. In both cases, the citizenry rose up against Beijing’s using the forging of closer economic and trade ties in the context of its political and economic rise to penetrate and dissipate Hong Kong’s right to autonomy, as well as Taiwan’s independence, autonomy and sovereign rights.” (*Taipei Times*, [October 13, 2014](#))

However, unlike in South Korea, the Taiwanese debate also was closely tied to statements of political leaders, such as President Ma Ying-jeou (KMT) and members of the then-opposition party DPP. The political actors expressed their support for the Umbrella Movement. In *The China Post* the statements by President Ma are presented in detail:

“Ma, however, stated that if the mainland authorities do not handle the Hong Kong people’s appeals with a delicate touch, progress in improvements in cross-strait relations would be hampered by further alienation and widening gaps in perception.” (*The China Post*, [October 1, 2014](#))

The *Taipei Times* on the other hand provided more space for DPP members to make statements:

“‘How could Ma be so shameless as to say he is proud of Taiwan’s democracy?’ Huang said. ‘It is ironic and pretentious for him to call on China to democratize, as he often caters to China at the expense of Taiwan’s democracy and the rule of law, while Hong Kongers are bravely fighting against the Chinese Communist Party for democracy and freedom.’” (*Taipei Times*, [October 11, 2014](#))

¹Other examples are *Taipei Times* ([October 11, 2014](#)) and *Taipei Times* ([November 25, 2014](#))

Given that the *Taipei Times* is closely linked to the DPP (Rawnsley and Rawnsley 2004, p. 6), it is no surprise that the newspaper seizes the opportunity to provide space for the DPP and to draw parallels with the Sunflower Movement.

It shows that the Umbrella Movement was a relevant topic for Taiwan's domestic political debate. It was less heavily debated in the UK though, but the protests were contextualized with UK's special responsibilities towards Hong Kong and the 1984 Joint Declaration (Sparks 2015, p. 435-436)

“At the very least, Britain should act honourably: it has a moral and legal responsibility to Hong Kong. It did after all sign a treaty, back in 1984, that guaranteed Hong Kong's core values and way of life, including freedom of speech and assembly, until 2047.” (*The Guardian*, October 5, 2014)

7 CONCLUSIONS

Based on the relevance of mediated public diplomacy and mediated discourses, this thesis investigated the similarities and differences in the news coverage of the 2014 protests in Hong Kong in English news outlets from East Asia. The basis for investigating the patterns of reporting on the Umbrella Movement was a corpus of 2,197 documents from nine English language news outlets. The documents were analyzed with automated content analysis techniques for indicators relating to salience, framing, and quotation preferences. The performance of the automated techniques varied. Extraction of meta information such as the documents' word count worked accurately. Manually composed search strings for identifying the news agency and relevant documents had a good performance. The quotation attribution via Stanford's *CoreNLP* generated reasonable results, but it required a large amount of manual work for speaker disambiguation. Even though the sentiment dictionary by Rinker (2019) supposedly has a good performance and is said to be for general purpose classifications, it yielded statistically non-significant results. The keyword assisted LDA model performed moderately well and seems suitable for automated frame coding. *Latent Semantic Scaling* brought mixed results with difficult interpretability. The negative values matched the concept of violent behavior well, but positive values included too many terms that go beyond the concept of peacefulness. Given that the text model was not used like this before and the seed words were somewhat experimental, there is room for improvement for working with it.

The results suggest that geographic and political proximity have an influence on media salience of the protests. More importantly, each country exhibited a distinct framing pattern. Within a country, the newspapers did not differ too much from each other in terms of their reporting patterns. The newspapers from South Korea are an exception, they employed opposing patterns, but also gave the issue low attention. Unsurprisingly, Taiwanese newspapers were generally in favor of the protests, while the two Chinese outlets were skeptical, and the Hong Kong based newspapers had a tendency to be neutral. Favorable reporting patterns included a preference to quote protesters and to mention their cause often. Against the expectations, episodes of violence were highlighted by newspapers that used patterns that were favorable of the protesters. Sensationalism and sympathy seem to be reasonable answers to this characteristic. Chinese newspapers established a bias against the Umbrella Movement through a preference for quoting government sources and, unexpectedly, by avoiding emphasizing violent behaviors. Instead, they emphasized the rule of law and how the protesters defied it.

Latent effects of news agency copy could be found in mainland China and in Taiwan. The Chinese newspapers both relied on the state agency *Xinhua*. Additionally, *People's Daily* sourced content from *China Daily* and various other PRC based news outlets. The Taiwanese newspapers strongly relied on Western agencies such as *Associated Press* and *Agence France-Presse*. Especially in Taiwan this is suspected to have caused patterns of parallelism.

In terms of pluralism, this study suggests that South Korea appears to have a vibrant discourse within the conservative camp on foreign issues. This aligns with the South Korean's preference for democratic values that relate to freedom and liberty. The differences in the other countries were not that strong, which hints that these seem to have a more unified mediated discourse in English.

The manual inspection of the text model results together with the suggestions by the works of Sparks (2015) and Ha and Shin (2016) revealed that some newspapers contextualized the foreign issue in terms of their domestic affairs. Taiwanese newspapers exhibited this pattern in particular.

The results of this study are very limited in terms of their scope and coverage. The selection of English language newspapers had the advantage of leveraging existing text models and to seamlessly compare mediated discourses from different countries. However, investigating the discourses in the native languages could bring promising results in terms of exploring pluralism in the respective media systems. Especially in terms of the mediated debate around foreign affairs in democratic systems such as South Korea or Taiwan. Additionally, the Japanese point of view was not covered, mainly owed to the lack of available text data. Some of the text models that were used in this thesis are generally language independent such as *Latent Semantic Scaling* and *Keyword Assisted Latent Dirichlet Allocation*. However, tuning these models for multilingual compatibility is a challenging new frontier. It could also bring new possibilities for content analysis that spans across languages and media.

Furthermore, there is potential to enhance the analysis of the data that was collected and generated for this thesis. While the framing automated classification of the cause frame and foreign influence frame was rather successful, there is not enough clarity on the valence of these frames. Combining the results of a valence analysis with the results of the LDA classification could provide a more nuanced view. The manual inspection showed that frames were frequently embedded in quotes. A correlation analysis could also provide more insights in the relationship between quotes and framing as it was done by Watanabe (2017a, p. 237). Also, a more complex analysis could potentially reveal how the foreign influence framing of the Umbrella Movement disseminated across news outlets.

Moving forward, the Umbrella Movement turned out to be the precursor for the protests that emerged in Hong Kong in 2019. Five years after the occupation, protesters dragged a large yellow banner to the streets in Admiralty stating: "We are back" (*Hong Kong Free Press*, September 28, 2019). Comparing the findings on the 2014 protests with the mediated discourse of the recent protests could bring new insights.

BIBLIOGRAPHY

- Antoniades, Andreas. 2012. "At the Eye of the Cyclone: The Greek Crisis in Global Media." In *The Konstantinos Karamanlis Institute for Democracy Series on European and International Affairs*, 11–25. Springer Berlin Heidelberg, December.
- Aschenbrenner, Andreas. 2019. "Emotionserkennung bei Nachrichtenkommentaren mittels Convolutional Neural Networks und Label Propagationsverfahren," Ludwig-Maximilians-Universität München, April. <http://nbn-resolving.de/urn:nbn:de:bvb:19-252950>.
- Asian Barometer Survey. 2017. (Wave 3: 13 Countries in East Asia 2010-2012). <http://asianbarometer.org/>.
- . 2018. (Wave 4: 14 Countries in East Asia 2014-2016). <http://asianbarometer.org/>.
- Backhaus, Klaus, Bernd Erichson, Wulff Plinke, and Rolf Weiber. 2016. *Multivariate Analysemethoden*. 14th ed. Springer Berlin Heidelberg.
- Baden, Christian, and Katsiaryna Stalpouskaya. 2015. "Common methodological framework: Content analysis. A mixed-methods strategy for comparatively, diachronically analyzing conflict discourse." *INFOCORE Working Paper*, nos. 2015/10, http://www.infocore.eu/wp-content/uploads/2016/02/Methodological-Paper-MWG-CA_final.pdf.
- Balandin, S. I., A. Filchenkov, L. M. Pivovarov, and J. Zizka., eds. 2016. *Proceedings of the Artificial Intelligence and Natural Language AINL FRUCT 2016 Conference*. Saint-Petersburg, Russia, November 10, 2016.
- Belle, Douglas A. Van. 2010. "Media Agenda Setting and Donor Aid." In *Public Sentinel*, edited by Pippa Norris, 85–108. The World Bank.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software* 3 (30): 774–778. <https://quanteda.io>.
- Blei, David M. 2012. "Probabilistic topic models." *Communications of the ACM* 55 (4): 77–84.
- Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent dirichlet allocation." *Journal of machine Learning research* 3 (Jan): 993–1022.

- Bolton, Kingsley, ed. 2002. *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press.
- Boumans, Jelle W., and Damian Trilling. 2015. "Taking Stock of the Toolkit." *Digital Journalism* 4 (1): 8–23.
- Box, G.E.P. 1979. "Robustness in the Strategy of Scientific Model Building." In *Robustness in Statistics*, 201–236. Elsevier.
- Boyle, Michael P. 2015. "A Tale of Two Sites." *Contention* 3, no. 1 (January): 17–35.
- Boyle, Michael P., and Douglas M. McLeod. 2018. "News Framing and Social Protest. Toward a Comprehensive Model." In *Doing News Framing Analysis*, 2nd ed., edited by Paul D'Angelo, 295–319. New York: Routledge.
- Brown, David G., and Kyle Churchman. 2019. "Hong Kong Impacts Taiwan Elections." *Comparative Connections* 21 (2): 65–72.
- Brüggemann, Michael, and Paul D'Angelo. 2018. "Defragmenting News Framing Research." In *Doing News Framing Analysis*, 2nd ed., edited by Paul D'Angelo, 90–111. New York: Routledge.
- Burscher, Björn, Daan Odijk, Rens Vliegthart, Maarten de Rijke, and Claes H. de Vreese. 2014. "Teaching the Computer to Code Frames in News: Comparing Two Supervised Machine Learning Approaches to Frame Analysis." *Communication Methods and Measures* 8, no. 3 (July): 190–206.
- Cao, Juan, Tian Xia, Jintao Li, Yongdong Zhang, and Sheng Tang. 2009. "A density-based method for adaptive LDA model selection." *Neurocomputing* 72, nos. 7–9 (March): 1775–1781.
- Carragee, Kevin M., and Wim Roefs. 2004. "The Neglect of Power in Recent Framing Research." *Journal of Communication* 54 (2): 214–233.
- Chan, Joseph M., and Chin-Chuan Lee. 1989. "Shifting Journalistic Paradigms: Editorial Stance and Political Transition in Hong Kong." *The China Quarterly* 117 (March): 97–117.
- Chan, Yuen-ying. 2002. "The English-language media in Hong Kong." In *Hong Kong English: autonomy and creativity*, edited by Kingsley Bolton, 101–115. Hong Kong: Hong Kong University Press.
- Chang, Tsan-Kuo, and Fen Lin. 2014. "From propaganda to public diplomacy: Assessing China's international practice and its image, 1950–2009." *Public Relations Review* 40, no. 3 (September): 450–458.
- Chang, Yu-tzung., Yunhan. Zhu, and Chong-min Pak. 2007. "Authoritarian Nostalgia in Asia." *Journal of Democracy* 18 (3): 66–80.

- Chen, Danqi, and Christopher Manning. 2014. "A Fast and Accurate Dependency Parser using Neural Networks." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 740–750. Doha, Qatar: Association for Computational Linguistics, October. <https://www.aclweb.org/anthology/D14-1082>.
- Cheng, Zhuqing, Guy J. Golan, and Spiro Kioussis. 2015. "The Second-Level Agenda-Building Function of the Xinhua News Agency." *Journalism Practice* 10, no. 6 (July): 744–762.
- Chiao, Chih-Hai, and Ju Chuan Huang. 2017. "Hidden Messages in News: A Comparison of English-language and Chinese-language Newspapers that report Sunflower Student Movement in Taiwan." *Critical Approaches to Discourse Analysis across Disciplines* 9 (1): 27–44.
- China Daily*. 2014. "Academic backs government's leadership of HK political reforms." August 30, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/china/2014-08/30/content_18516985.htm.
- . 2014. "Editorial: Get HKSAR back on track." November 28, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/opinion/2014-11/28/content_18990476.htm.
- . 2014. "HK's reform must follow constitutional framework." By Leung Mei-Fun. October 6, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/hkedition/2014-10/06/content_18699467.htm.
- . 2014. "MFA reiterates opposition to foreign interference in HK." By Joseph Li. October 1, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/hkedition/2014-10/01/content_18689876.htm.
- . 2014. "Obama and Abe have shown their ignorance of history." By Hon-Kuen Ho. May 12, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/hkedition/2014-05/12/content_17499482.htm.
- . 2014. "'Occupy Central' is a gamble which has failed." By Leung Kwok-leung. September 30, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/hkedition/2014-09/30/content_18684427.htm.
- . 2014. "Our youth must not waste a bright future." By Raymond So. October 16, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/hkedition/2014-10/16/content_18745733.htm.
- . 2014. "Protests may breed youth angst in Hong Kong SAR." By Harry Ong. October 27, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/hkedition/2014-10/27/content_18805103.htm.
- . 2014. "West slammed for double standards over Hong Kong." October 23, 2014. Accessed March 26, 2020. http://www.chinadaily.com.cn/world/2014-10/23/content_18791000.htm.

- Chosun Ilbo. 2014. "Hong Kong Leader Sees 'Foreign' Hand." October 20, 2014. Accessed March 22, 2020. http://english.chosun.com/site/data/html_dir/2014/10/20/2014102000582.html.
- Chu, Yun-han, and Min-hua Huang. 2010. "Solving an Asian Puzzle." *Journal of Democracy* 21 (4): 114–122.
- Chu, Yun-han, Liu Kang, and Min-hua Huang. 2014. "How East Asians View the Rise of China." *Journal of Contemporary China* 24, no. 93 (October): 398–420.
- Copper, John F. 2020. "Taiwan's 2020 Presidential, Vice Presidential and Legislative Election: Testing Election Theories." *East Asian Policy* 12, no. 02 (April): 45–56.
- D'Angelo, Paul. 2018a. "A Typology of Frames in News Framing Analysis." In *Doing News Framing Analysis*, 2nd ed., edited by Paul D'Angelo, xxiii–xl. New York: Routledge.
- , ed. 2018b. *Doing News Framing Analysis: Empirical and Theoretical Perspectives*. 2nd ed. New York: Routledge.
- Deveaud, Romain, Eric SanJuan, and Patrice Bellot. 2014. "Accurate and effective latent concept modeling for ad hoc information retrieval." *Document numérique* 17, no. 1 (April): 61–84.
- Donsbach, Wolfgang, ed. 2008. *The International Encyclopedia of Communication*. John Wiley & Sons, Ltd.
- Du, Ying Roselyn. 2015. "Same Events, Different Stories." *Journalism & Mass Communication Quarterly* 93, no. 1 (October): 99–117.
- Eberl, Jakob-Moritz, Hajo G. Boomgaarden, and Markus Wagner. 2015. "One Bias Fits All? Three Types of Media Bias and Their Effects on Party Preferences." *Communication Research*, 1125–1148.
- Elmasry, Mohamad H., and Vidhi Chaudhri. n.d. "The Press as Agent of Cultural Repair: A Textual Analysis of News Coverage of the Virginia Tech Shootings." In *The Handbook of Crisis Communication*, 141–158. Wiley-Blackwell.
- Entman, Robert M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43 (4): 51–58.
- . 2008. "Theorizing Mediated Public Diplomacy: The U.S. Case." *The International Journal of Press/Politics* 13, no. 2 (March): 87–102.
- Eshima, Shusei, Kosuke Imai, and Tomoya Sasaki. 2020. "Keyword Assisted Topic Models" (April 13, 2020). arXiv: 2004.05964 [CS.CL].
- Feng, William Dezheng. 2017. "Ideological dissonances among Chinese-language newspapers in Hong Kong: A corpus-based analysis of reports on the Occupy Central Movement." *Discourse & Communication* 11, no. 6 (September): 549–566.

- Foot, Rosemary, and Evelyn Goh. 2018. "The International Relations of East Asia: A New Research Prospectus." *International Studies Review* 21, no. 3 (April): 398–423.
- Griffiths, T. L., and M. Steyvers. 2004. "Finding scientific topics." *Proceedings of the National Academy of Sciences* 101, no. Supplement 1 (February): 5228–5235.
- Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–297.
- Grusky, Max, Mor Naaman, and Yoav Artzi. 2018. "NEWSROOM: A Dataset of 1.3 Million Summaries with Diverse Extractive Strategies." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 708–719. New Orleans, Louisiana: Association for Computational Linguistics, June. <http://aclweb.org/anthology/N18-1065>.
- Gunter, Berrie. 2002. "The quantitative research process." In *A handbook of media and communication research*, edited by Klaus Bruhn Jensen, 209–234. London New York: Routledge.
- Ha, Jae Sik. 2015. "Arab Spring is Wake-Up Call for Kim Jong-Il." *Journalism Practice* 9, no. 5 (January): 741–759.
- Ha, Jae Sik, and Donghee Shin. 2016. "Framing the Arab Spring: Partisanship in the news stories of Korean Newspapers." *International Communication Gazette* 78, no. 6 (July): 536–556.
- . 2019. "Portraying China as an alternative to U. S. Hegemony: The China daily's framing of the arab spring." *Atlantic Journal of Communication* 27, no. 3 (April): 200–215.
- Han, Gang (kevin). 2007. "Mainland China Frames Taiwan: How China's News Websites Covered Taiwan's 2004 Presidential Election." *Asian Journal of Communication* 17, no. 1 (March): 40–57.
- Harlow, Summer, and Thomas J. Johnson. 2011. "Overthrowing the Protest Paradigm? How The New York Times, Global Voices and Twitter Covered the Egyptian Revolution." *International Journal of Communication* 5:1359–1374.
- Hartig, Falk. 2016. "How China Understands Public Diplomacy: The Importance of National Image for National Interests." *International Studies Review* (June): 655–680.
- Heidenreich, Tobias, Jakob-Moritz Eberl, Fabienne Lind, and Hajo Boomgaarden. 2019. "Political migration discourses on social media: a comparative perspective on visibility and sentiment across political Facebook accounts in Europe." *Journal of Ethnic and Migration Studies* 46, no. 7 (October): 1261–1280.

- Hilton, Anthony, and Richard A. Armstrong. 2006. "Statnote 6: post-hoc ANOVA tests." *Microbiologist* 7 (3): 34–36.
- Ho, Ming-sho. 2019. *Challenging Beijing's mandate of heaven: Taiwan's Sunflower Movement and Hong Kong's Umbrella Movement*. Philadelphia: Temple University Press.
- Hong Kong Free Press. 2019. "'We are back': Hong Kong protesters recreate pro-democracy 'Lennon Wall' on 5th anniversary of Umbrella Movement." By Jennifer Creery. September 28, 2019. Accessed December 8, 2020. <https://hongkongfp.com/2019/09/28/back-hong-kong-protesters-recreate-pro-democracy-lennon-wall-5th-anniversary-umbrella-movement/>.
- Jacobi, Carina, Wouter van Atteveldt, and Kasper Welbers. 2015. "Quantitative analysis of large amounts of journalistic texts using topic modelling." *Digital Journalism* 4, no. 1 (October): 89–106.
- Jensen, Klaus Bruhn, ed. 2002a. *A handbook of media and communication research: qualitative and quantitative methodologies*. London New York: Routledge.
- . 2002b. "Introduction: The state of convergence in media and communication research." In *A handbook of media and communication research*, edited by Klaus Bruhn Jensen, 1–11. London New York: Routledge.
- . 2002c. "The complementarity of qualitative and quantitative methodologies in media and communication research." In *A handbook of media and communication research*, edited by Klaus Bruhn Jensen, 254–272. London New York: Routledge.
- Karyotakis, Minos-Athanasios, Nikos Panagiotou, Nikos Antonopoulos, and Matina Kiourexidou. 2017. "Digital Media Framing of the Egyptian Arab Spring: Comparing Al Jazeera, BBC and China Daily." *Studies in Media and Communication* 5, no. 2 (October): 66–75.
- Kilgo, Danielle K., Summer Harlow, Victor García-Perdomo, and Ramón Salaverría. 2018. "From #Ferguson to #Ayotzinapa: Analyzing Differences in Domestic and Foreign Protest News Shared on Social Media." *Mass Communication and Society* 21, no. 5 (June): 606–630.
- Kim, Kisun, and Saif Shahin. 2020. "Ideological parallelism: toward a transnational understanding of the protest paradigm." *Social Movement Studies* 19, no. 4 (October): 391–407.
- Krippendorff, Klaus. 2004. *Content analysis: an introduction to its methodology*. Thousand Oaks, Calif: Sage.
- Lai, Christina. 2017. "Acting one way and talking another: China's coercive economic diplomacy in East Asia and beyond." *The Pacific Review* 31, no. 2 (July): 169–187.
- Lee, Francis L. F., and Joseph M. Chan. 2018. *Media and Protest Logics in the Digital Era*. Oxford University Press.

- Lee, Hsiao-wen. 2015. "From control to competition: A comparative study of the party press and popular press." In *Routledge Handbook of Chinese Media*, edited by Gary D. Rawnsley, 115–130.
- Lee, Jae-won. 2011. "South Korea: Media System." In *The International Encyclopedia of Communication*, edited by Wolfgang Donsbach. John Wiley & Sons, Ltd.
- Lee, Jung-Eun. 2012. "Micro-Dynamics of Protests: The Political and Cultural Conditions for Anti-U.S. Beef Protests in South Korea." *Sociological Perspectives* 55, no. 3 (August): 399–420.
- Loh, Christine. 2010. *Underground Front: The Chinese Communist Party in Hong Kong*. Hong Kong University Press.
- Lu, Jie. 2012. "Democratic Conceptions in East Asian Societies: A Contextualized Analysis." *Asian Barometer Working Paper Series*, no. 56, <http://asianbarometer.org/publications/cadc6e97339be04348ac25e819d93aab.pdf>.
- Maier, Daniel, A. Walldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, et al. 2018. "Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology." *Communication Methods and Measures* 12, nos. 2-3 (February): 93–118.
- McCluskey, Michael, Susan E. Stein, Michael P. Boyle, and Douglas M. McLeod. 2009. "Community Structure and Social Protest: Influences on Newspaper Coverage." *Mass Communication and Society* 12, no. 3 (June): 353–371.
- McLeod, Douglas M., and Benjamin H. Detenber. 1999. "Framing Effects of Television News Coverage of Social Protest." *Journal of Communication* 49, no. 3 (September): 3–23.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space" (January 16, 2013). arXiv: [1301.3781](https://arxiv.org/abs/1301.3781) [cs.CL].
- Molen, Sherri L. Ter. 2011. "The 2008 US presidential campaign as represented in the online edition of The Korea Times." *Media Asia* 38 (2): 67–77.
- Murdock, Graham. 2002. "Media, culture and modern times: social science investigations." In *A handbook of media and communication research*, edited by Klaus Bruhn Jensen, 40–61. London New York: Routledge.
- Murzintcev, Nikita. 2020. *ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters*. April 21, 2020. Accessed August 13, 2020. <https://CRAN.R-project.org/package=ldatuning>.
- Muzny, Grace, Michael Fang, Angel Chang, and Dan Jurafsky. 2017. "A Two-stage Sieve Approach for Quote Attribution." In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 460–470. Valencia, Spain: Association for Computational Linguistics, April. <https://www.aclweb.org/anthology/E17-1044>.

- Naldi, Maurizio. 2019. "A review of sentiment computation methods with R packages" (January 24, 2019). arXiv: [1901.08319](https://arxiv.org/abs/1901.08319) [cs.CL].
- Norris, Pippa, ed. 2010. *Public Sentinel: News Media & Government Reform*. The World Bank.
- O'Neill, Deirdre, and Tony Harcup. 2019. "News Values and News Selection." In *The Handbook of Journalism Studies*, edited by Karin Wahl-Jorgensen and Thomas Hanitzsch, 213–228. Routledge.
- Ojala, Markus, and Mervi Pantti. 2017. "Naturalising the new cold war: The geopolitics of framing the Ukrainian conflict in four European newspapers." *Global Media and Communication* 13, no. 1 (February): 41–56.
- Pashakhin, Sergei. 2016. "Topic Modeling for Frame Analysis of News Media." In *Proceedings of the Artificial Intelligence and Natural Language AINL FRUCT 2016 Conference*, 103–105.
- People's Daily*. 2014. "China's central authorities fully support HK gov't: CY Leung." October 20, 2014. Accessed February 15, 2020. <http://en.people.cn/n/2014/1020/c90785-8796824.html>.
- . 2014. "Reform in Hong Kong must follow the Basic Law: article." By Wang Xin and Liang Jun. July 10, 2014. Accessed February 15, 2020. <http://en.people.cn/n/2014/0710/c90785-8753487.html>.
- Rawnsley, Gary D. 2015a. "Chinese international broadcasting, public diplomacy and soft power." In *Routledge Handbook of Chinese Media*, edited by Gary D. Rawnsley, 460–475.
- , ed. 2015b. *Routledge Handbook of Chinese Media*. Routledge, April.
- Rawnsley, Gary D., and Ming-Yeh T. Rawnsley. 2004. "Media reform since 1987." *China Perspectives* 2004, no. 6 (December): 1–15.
- Reuters*. 2019. "Hundreds of thousands march in Hong Kong to protest China extradition bill." By James Pomfret and Anne Marie Roantree. June 9, 2019. <https://www.reuters.com/article/us-hongkong-extradition-march-idUSKCN1TA05L>.
- Al-Rfou, Rami, Bryan Perozzi, and Steven Skiena. 2013. "Polyglot: Distributed Word Representations for Multilingual NLP" (July 5, 2013). arXiv: [1307.1662](https://arxiv.org/abs/1307.1662) [cs.CL].
- Rinker, Tyler W. 2019. *sentimentr: Calculate Text Polarity Sentiment*. Version 2.7.1. Buffalo, New York. <http://github.com/trinker/sentimentr>.
- Ryfe, David. 2019. "Journalism and Democracy." In *The Handbook of Journalism Studies*, edited by Karin Wahl-Jorgensen and Thomas Hanitzsch, 293–306. Routledge.
- Schröder, Kim Christian. 2002. "Discourses of fact." In *A handbook of media and communication research*, edited by Klaus Bruhn Jensen, 98–116. London New York: Routledge.

- SCMP. 2014. "Doomed' protests will not change Beijing's decision on political reform, warns Chinese state paper." By James Griffiths. September 29, 2014. Accessed February 4, 2020. <https://www.scmp.com/news/hong-kong/article/1603840/doomed-protests-will-not-change-beijings-decision-political-reform>.
- . 2014. "Fears over 'radicals' as protest violence increases, but sources say Beijing won't be embarrassed into action." By Alice Wu. October 20, 2014. Accessed February 4, 2020. <https://www.scmp.com/news/hong-kong/article/1620188/fears-over-radicals-protest-violence-increases-sources-say-beijing>.
- . 2014. "Hong Kong is strong enough to survive Occupy Central." By Alex Lo. June 26, 2014. Accessed February 4, 2020. <https://www.scmp.com/comment/insight-opinion/article/1540490/hong-kong-strong-enough-survive-occupy-central>.
- . 2014. "Illustrate your opinion: artists draw in support of Hong Kong protesters." October 2, 2014. Accessed February 4, 2020. <https://www.scmp.com/news/hong-kong/article/1607587/illustrate-your-opinion-artists-draw-support-hong-kong-protesters>.
- . 2014. "Loss of confidence a hazard of occupation." By Richard Harris. December 4, 2014. Accessed February 4, 2020. <https://www.scmp.com/comment/insight-opinion/article/1655427/loss-confidence-hazard-occupation>.
- . 2014. "Malaysian couple to hang for murder of Indonesian maid who died of starvation." March 7, 2014. Accessed February 4, 2020. <https://www.scmp.com/news/asia/article/1442586/malaysian-couple-hang-murder-indonesian-maid-who-died-starvation>.
- . 2014. "Our youth are special ... but no more so than others before them." By Alice Wu. November 10, 2014. Accessed February 4, 2020. <https://www.scmp.com/comment/insight-opinion/article/1634357/our-youth-are-special-no-more-so-others-them>.
- . 2014. "The local heroes who set an example for us all." October 19, 2014. Accessed February 4, 2020. <https://www.scmp.com/comment/insight-opinion/article/1619526/local-heroes-who-set-example-us-all>.
- Semetko, Holli A., and Patti M. Valkenburg. 2000. "Framing European politics: A Content Analysis of Press and Television News." *Journal of Communication* 50, no. 2 (June): 93–109.
- Shahin, Saif, Pei Zheng, Heloisa Aruth Sturm, and Deepa Fadnis. 2016. "Protesting the Paradigm." *The International Journal of Press/Politics* 21, no. 2 (February): 143–164.
- Sklias, Pantelis, and Nikolaos Tzifakis, eds. 2013. *Greece's Horizons*. Springer Berlin Heidelberg.

- Sparks, Colin. 2015. "Business as usual: the UK national daily press and the Occupy Central movement." *Chinese Journal of Communication* 8, no. 4 (October): 429–446.
- Sullivan, Jonathan, and Chun-Yi Lee, eds. 2018. *A New Era in Democratic Taiwan*. Routledge.
- Sullivan, Jonathan, Ming-yeh T. Rawnsley, Chien-san Feng, and James Smyth. 2018. "The Media in Democratic Taiwan." In *A New Era in Democratic Taiwan*, 104–122.
- Tai, Benny. 2013. "公民抗命的最大殺傷力武器 (The Most Lethal Weapon of Civil Disobedience)" [in Chinese]. Translation by author available at: <https://bennytai.github.io/HongKongReflections/思考香港6/2-The%20Most%20Lethal%20Weapon%20of%20Civil%20Disobedience.html>, 信報財經新聞 (*Hong Kong Economic Journal*) (January 16, 2013). Accessed May 14, 2020. <https://www1.hkej.com/dailynews/article/id/654855/>.
- Taipei Times. 2014. "HK protests a peaceful revolution." By Paul Lin. October 11, 2014. Accessed March 3, 2020. <https://www.taipeitimes.com/News/editorials/archives/2014/10/11/2003601789>.
- . 2014. "Hong Kong culls 20,000 birds amid H7N9 flu scare." January 29, 2014. Accessed March 3, 2020. <https://taipeitimes.com/News/world/archives/2014/01/29/2003582428>.
- . 2014. "Lessons from 'umbrella revolution'." October 8, 2014. Accessed March 3, 2020. <https://www.taipeitimes.com/News/editorials/archives/2014/10/08/2003601553>.
- . 2014. "'One country, two systems' model is inappropriate for Taiwan: analyst." By William Lowther. November 25, 2014. Accessed March 3, 2020. <https://www.taipeitimes.com/News/taiwan/archives/2014/11/25/2003605271>.
- . 2014. "Taiwanese, HK youths protecting democracy." By Chang Ching-yun. October 13, 2014. Accessed March 3, 2020. <https://www.taipeitimes.com/News/editorials/archives/2014/10/13/2003601931>.
- . 2014. "Tsai Ing-wen, Sunflower leader slam Ma comments." By Loa Iok-sin. October 11, 2014. Accessed March 3, 2020. <https://www.taipeitimes.com/News/front/archives/2014/10/11/2003601796>.
- . 2014. "World watches Beijing in Hong Kong." October 7, 2014. Accessed March 3, 2020. <https://taipeitimes.com/News/world/archives/2014/10/07/2003601506>.
- Templeman, Kharis. 2020. "How Taiwan Stands Up to China." *Journal of Democracy* 31 (3): 85–99.
- The China Post. 2014. "2014 has been a worrisome year." December 30, 2014. Accessed February 27, 2020. <https://chinapost.nownews.com/20141230-60050>.

- The China Post*. 2014. "China takes risk with anti-US fervor." By Frank Ching. October 15, 2014. Accessed February 27, 2020. <https://chinapost.nownews.com/20141015-71144>.
- . 2014. "Hong Kong officials, protesters hold 1st talks." By Aaron Tam and Dennis Chong. October 22, 2014. Accessed February 27, 2020. <https://chinapost.nownews.com/20141022-70479>.
- . 2014. "Hong Kong students plan pro-democracy protest." By Aaron Tam and Dennis Chong. September 5, 2014. Accessed February 27, 2020. <http://www.taipeitimes.com/News/world/archives/2014/09/05/2003599072>.
- . 2014. "Ruling and opposition parties express stances on Hong Kong protests." By Ted Chen. October 1, 2014. <https://chinapost.nownews.com/20141001-49346>.
- The Guardian*. 2014. "China: a great power we can't trust." By Isabel Hilton. July 1, 2014. Accessed March 25, 2020. <https://www.theguardian.com/commentisfree/2014/jul/01/china-promise-hong-kong-liberties-power>.
- . 2014. "Hong Kong – betrayed by China. And abandoned by the British." By Anson Chan. October 5, 2014. Accessed March 25, 2020. <https://www.theguardian.com/commentisfree/2014/oct/05/hong-kong-protests-betrayed-by-china-abandoned-by-britain>.
- . 2014. "Hong Kong police detain activists during democracy protests." September 5, 2014. Accessed March 25, 2020. <https://www.theguardian.com/world/2014/sep/02/hong-kong-police-democracy-protests-china>.
- . 2014. "Hong Kong: voices from the 'umbrella revolution'." By Tania Branigan. September 30, 2014. Accessed March 25, 2020. <https://www.theguardian.com/world/2014/sep/30/-sp-hong-kong-umbrella-protest-demonstrators-voices>.
- . 2014. "Leading Hong Kong activist accuses Cameron of selling out campaigners." By Luke Harding and Richard Norton-Taylor. September 30, 2014. Accessed March 25, 2020. <https://www.theguardian.com/world/2014/sep/30/hong-kong-accuse-cameron-selling-out-china>.
- The Korea Times*. 2014. "China's concern on Western interference in Hong Kong." By Frank Ching. October 21, 2014. Accessed March 29, 2020. http://www.koreatimes.co.kr/www/nation/2014/10/113_166693.html.
- . 2014. "Hong Kong's institutions of democracy." By Henrique Schneider. October 21, 2014. Accessed March 29, 2020. http://www.koreatimes.co.kr/www/opinion/2014/10/137_166703.html.
- . 2014. "Hong Kong's isolated crusade." By Donald Kirk. October 23, 2014. Accessed March 29, 2020. http://www.koreatimes.co.kr/www/opinion/2014/10/137_166845.html.

- The Korea Times*. 2014. "Protesting for democracy." By Donald Kirk. October 2, 2014. Accessed March 29, 2020. http://www.koreatimes.co.kr/www/opinion/2014/10/137_165636.html.
- The New York Times*. 2018. "A Hong Kong Newspaper on a Mission to Promote China's Soft Power." By Javier C. Hernández. March 31, 2018. Accessed May 22, 2020. <https://www.nytimes.com/2018/03/31/world/asia/south-china-morning-post-hong-kong-alibaba.html>.
- The Standard*. 2014. "Beijing 'will be pleased' by scrapping of vote." October 27, 2014. Accessed March 30, 2020. <https://www.thestandard.com.hk/breaking-news/section/4/54404/Beijing-'will-be-pleased'-by-scrapping-of-vote>.
- . 2014. "Democracy protests 'undemocratic': China party paper." October 6, 2014. Accessed March 30, 2020. <https://www.thestandard.com.hk/breaking-news/section/4/53920/Democracy-protests-'undemocratic':-China-party-paper>.
- . 2014. "SARS-like fall in shop sales." By Kenneth Lau and Hilary Wong. October 3, 2014. Accessed March 30, 2020. <https://www.thestandard.com.hk/section-news/section/11/150026/SARS-like-fall-in-shop-sales>.
- . 2014. "Set aside differences for sake of suffrage, says Li." September 2, 2014. Accessed March 30, 2020. <https://www.thestandard.com.hk/section-news/section/11/149040/Set-aside-differences-for-sake-of-suffrage,-says-Li>.
- Trilling, Damian, and Jeroen G. F. Jonkman. 2018. "Scaling up Content Analysis." *Communication Methods and Measures* 12, nos. 2-3 (March): 158–174.
- Tuchman, Gaye. 2002. "The production of news." In *A handbook of media and communication research*, edited by Klaus Bruhn Jensen, 78–90. London New York: Routledge.
- VOA News. 2014. "香港佔領運動清場前市民反應 [Public reactions before the eviction of the Hong Kong Occupy Movement]" [in Chinese]. December 10, 2014. Accessed December 8, 2020. <https://www.voacantonese.com/a/hong-kong-protest-sites-to-be-cleared-citizens-reactions/2552004.html>.
- Voice of America Office of Public Relations. n.d. "Mission and Values." Accessed November 16, 2020. <https://www.insidevoa.com/p/5831.html>.
- Vreese, Claes de, and Hajo Boomgaarden. 2003. "Valenced news frames and public support for the EU." *Communications* 28, no. 4 (January): 361–381.
- Wahl-Jorgensen, Karin, and Thomas Hanitzsch, eds. 2019. *The Handbook of Journalism Studies*. Routledge.
- Wang, Shaojung Sharon, and Junhao Hong. 2011. "Voice of America in the post-Cold War era: Opportunities and challenges to external media services via new information and communication technology." *International Communication Gazette* 73, no. 4 (June): 343–358.

- Wanta, Wayne, Guy Golan, and Cheolhan Lee. 2004. "Agenda Setting and International News: Media Influence on Public Perceptions of Foreign Nations." *Journalism & Mass Communication Quarterly* 81 (2): 364–377.
- Watanabe, Kohei. 2017a. "Measuring news bias: Russia's official news agency ITAR-TASS' coverage of the Ukraine crisis." *European Journal of Communication* 32, no. 3 (March): 224–241.
- . 2017b. "Newsmap." *Digital Journalism* 6, no. 3 (March): 294–309.
- . 2017c. "Online Appendix: Measuring news bias: Russia's official news agency ITAR-TASS' coverage of the Ukraine crisis." *European Journal of Communication* 32, no. 3 (March). Accessed July 14, 2020. <https://blog.koheiw.net/wp-content/uploads/2017/02/Measuring-news-bias-06-appendix.pdf>.
- . 2020. "Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages." *Communication Methods and Measures* (November): 1–22.
- Watanabe, Kohei, and Yuan Zhou. 2020. "Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches." *Social Science Computer Review* (February): 1–21.
- Welbers, Kasper, Wouter Van Atteveldt, and Kenneth Benoit. 2017. "Text Analysis in R." *Communication Methods and Measures* 11 (4): 245–265.
- Wimmer, Roger, and Joseph R. Dominick. 2011. *Mass media research: An Introduction*. 9th ed. Boston, Mass: Cengage-Wadsworth.
- Wu, Jack Chen-chia, Mark Weatherall, and Yu-tzung Chang. 2012. "How Taiwanese Citizens View Democracy: Change and Continuity in Democratic Attitudes and Values in Taiwan's Democratic Consolidation." *Asian Barometer Working Paper Series*, no. 55, <http://asianbarometer.org/publications/ea7e6be8e81526a6ab2738e30c561c0e.pdf>.
- Zhang, Li. 2010. "The Rise of China: media perception and implications for international politics." *Journal of Contemporary China* 19 (64): 233–254.
- Zhou, Shuhua. 2015. "China: Media System." In *The International Encyclopedia of Communication*, edited by Wolfgang Donsbach. John Wiley & Sons, Ltd.

MANUAL CODING RULES

The depiction of violent or peaceful behavior is coded on the document level. The intention is to capture the overall impression of the document. Each document is classified on a scale from 1 to 5:

- 1: document overwhelmingly contains depictions of violence.
- 2: document depicts violence. Including the use of physical force.
- 3: document is balanced (equal depictions of violence and peaceful behaviors). Or complete absence of violence or peacefulness in the document.
- 4: document mentions peacefulness. Including dialogue as means to settle conflicts.
- 5: document overwhelmingly includes descriptions of peaceful behaviors.

TABLES

Table 1: LSS Post-hoc test results territory comparison (Games-Howell)

Territory Pair	Difference	Lower bound	Upper bound	<i>t</i>	<i>df</i>	<i>p</i>
China-Hong Kong	0.515	0.372	0.657	9.875	840.421	0.000
China-Taiwan	0.416	0.224	0.609	5.92	677.692	0.000
China-UK	0.401	0.157	0.645	4.533	150.545	0.000
Taiwan-Hong Kong	0.098	-0.07	0.267	1.599	508.526	0.499
S. Korea-Hong Kong	0.269	-0.231	0.77	1.538	38.962	0.545
UK-Hong Kong	0.114	-0.112	0.34	1.396	111.499	0.632
China-S. Korea	0.246	-0.262	0.754	1.378	41.945	0.645
Taiwan-S. Korea	-0.171	-0.686	0.344	0.942	44.789	0.879
UK-S. Korea	-0.155	-0.689	0.379	0.821	52.17	0.923
UK-Taiwan	0.016	-0.244	0.275	0.165	185.788	1.000

Table 2: LSS Post-hoc test results newspaper comparison (Games-Howell)

Newspaper Pair	Difference	Lower bound	Upper bound	<i>t</i>	<i>df</i>	<i>p</i>
China Post-China Daily	-0.45	-0.76	-0.15	4.63	245.86	0.000
SCMP-China Daily	-0.51	-0.68	-0.33	8.87	744.35	0.000
The Standard-China Daily	-0.49	-0.72	-0.27	6.82	738.99	0.000
SCMP-People's Daily	-0.58	-0.98	-0.19	4.66	86.19	0.000
Taipei Times-China Daily	-0.36	-0.63	-0.10	4.27	333.97	0.001
The Guardian-China Daily	-0.39	-0.67	-0.11	4.32	159.50	0.001
The Standard-People's Daily	-0.57	-0.99	-0.15	4.30	108.61	0.001
People's Daily-China Post	0.53	0.06	0.99	3.57	151.70	0.014
The Guardian-People's Daily	-0.46	-0.91	-0.01	3.25	129.89	0.039
Taipei Times-People's Daily	-0.44	-0.88	0.00	3.13	128.41	0.054
SCMP-Korea Times	-0.42	-1.08	0.23	2.23	21.22	0.421
The Standard-Korea Times	-0.41	-1.07	0.25	2.11	23.61	0.491
Taipei Times-SCMP	0.15	-0.10	0.39	1.88	253.36	0.628
Korea Times-China Post	0.37	-0.32	1.06	1.80	29.23	0.680
The Guardian-Korea Times	-0.31	-0.98	0.37	1.51	27.15	0.840
People's Daily-Chosun Ilbo	0.49	-0.65	1.64	1.48	21.05	0.852
The Standard-Taipei Times	-0.13	-0.41	0.15	1.49	385.89	0.861
The Guardian-SCMP	0.12	-0.15	0.38	1.41	120.27	0.890
Taipei Times-Korea Times	-0.28	-0.95	0.40	1.39	26.00	0.891
Chosun Ilbo-China Daily	-0.42	-1.53	0.69	1.33	16.74	0.908
The Standard-The Guardian	-0.11	-0.40	0.19	1.11	187.78	0.972
Korea Times-Chosun Ilbo	0.34	-0.89	1.56	0.93	26.83	0.989
Taipei Times-China Post	0.09	-0.25	0.44	0.84	302.60	0.996
People's Daily-Korea Times	0.16	-0.57	0.89	0.71	38.30	0.998
Korea Times-China Daily	-0.08	-0.74	0.58	0.43	22.62	1.000
People's Daily-China Daily	0.08	-0.34	0.48	0.58	99.23	1.000
Chosun Ilbo-China Post	0.04	-1.09	1.16	0.11	18.54	1.000
SCMP-China Post	-0.05	-0.34	0.24	0.57	195.96	1.000
The Guardian-China Post	0.07	-0.30	0.43	0.57	232.04	1.000
The Standard-China Post	-0.04	-0.36	0.28	0.39	280.58	1.000
SCMP-Chosun Ilbo	-0.09	-1.20	1.02	0.28	16.35	1.000
Taipei Times-Chosun Ilbo	0.06	-1.06	1.18	0.18	17.67	1.000
The Guardian-Chosun Ilbo	0.03	-1.09	1.15	0.10	18.00	1.000
The Standard-Chosun Ilbo	-0.08	-1.19	1.04	0.24	17.02	1.000
The Standard-SCMP	0.01	-0.19	0.21	0.20	655.90	1.000
The Guardian-Taipei Times	-0.03	-0.35	0.30	0.27	217.53	1.000

Table 3: ABS Wave 3 Freedom & Liberty Index Post-hoc test results (Games-Howell)

Country Pair	Difference	Lower bound	Upper bound	<i>t</i>	<i>df</i>	<i>p</i>
Korea-Hong Kong	0.04	-0.05	0.13	1.01	2411.65	0.741
China-Hong Kong	-0.32	-0.39	-0.25	11.98	1624.48	<.001
Taiwan-Hong Kong	-0.15	-0.24	-0.07	4.63	2557.89	<.001
China-Korea	-0.36	-0.43	-0.29	13.43	1634.99	<.001
Taiwan-Korea	-0.19	-0.27	-0.1	5.76	2574.59	<.001
Taiwan-China	0.17	0.11	0.23	7.31	2356.41	<.001

Table 4: ABS Wave 4 Freedom & Liberty Index Post-hoc test results (Games-Howell)

Country Pair	Difference	Lower bound	Upper bound	<i>t</i>	<i>df</i>	<i>p</i>
Korea-Hong Kong	0.15	0.05	0.25	3.98	2407.76	<.001
China-Hong Kong	-0.35	-0.42	-0.27	11.93	1546.69	<.001
Taiwan-Hong Kong	-0.12	-0.21	-0.03	3.41	2473.29	0.004
China-Korea	-0.49	-0.57	-0.42	18.02	1569.82	<.001
Taiwan-Korea	-0.27	-0.35	-0.18	8.01	2544.96	<.001
Taiwan-China	0.23	0.17	0.29	9.72	2398.88	<.001

Table 5: Summary of quotation counts by actor groups (full table)

Territory	Newspaper	<i>n</i>	Bystanders		Counter Protest		HK Gov		Media		Other Governments		Others		PRC Gov		Protesters		Quotation		UK Gov		Unknown	
			<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
China	China Daily	376	6	1.60	2	0.53	71	18.88	6	1.60	10	2.66	9	2.39	24	6.38	52	13.83	283	75.27	10	2.66	241	64.10
	People's Daily	76	2	2.63	1	1.32	22	28.95	1	1.32	4	5.26	3	3.95	8	10.53	9	11.84	52	68.42	4	5.26	42	55.26
Hong Kong	SCMP	911	30	3.29	26	2.85	209	22.94	22	2.41	33	3.62	93	10.21	59	6.48	227	24.92	795	87.27	21	2.31	754	82.77
	The Standard	381	6	1.57	8	2.10	54	14.17	10	2.62	9	2.36	14	3.67	18	4.72	63	16.54	293	76.90	5	1.31	259	67.98
S. Korea	Chosun Ilbo	17	0	0.00	0	0.00	4	23.53	0	0.00	0	0.00	2	11.76	0	0.00	7	41.18	15	88.24	0	0.00	13	76.47
	Korea Times	21	0	0.00	0	0.00	2	9.52	0	0.00	0	0.00	2	9.52	3	14.29	1	4.76	19	90.48	0	0.00	17	80.95
Taiwan	China Post	152	1	0.66	0	0.00	21	13.82	4	2.63	4	2.63	13	8.55	13	8.55	38	25.00	104	68.42	2	1.32	88	57.89
	Taipei Times	175	2	1.14	1	0.57	45	25.71	4	2.29	5	2.86	18	10.29	31	17.71	62	35.43	169	96.57	2	1.14	156	89.14
UK	The Guardian	88	1	1.14	0	0.00	22	25.00	4	4.55	3	3.41	9	10.23	10	11.36	27	30.68	67	76.14	2	2.27	65	73.86
	Total	2197	48	2.18	38	1.73	450	20.48	51	2.32	68	3.10	163	7.42	166	7.56	486	22.12	1797	81.79	46	2.09	1635	74.42
<i>p</i>			0.236		0.034		< .001		0.830		0.825		< .001		< .001		< .001		< .001		0.464		< .001	