

# Automated quantification of Reuters news using a receiver operating characteristic curve analysis: The Western media image of China

Global Media and China

2017, Vol. 2(3–4) 251–268

© The Author(s) 2018

Reprints and permissions:

[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/2059436418754890

[journals.sagepub.com/home/gch](http://journals.sagepub.com/home/gch)

**Jukka Aukia, Juho Heimonen, Tapio Pahikkala  
and Tapio Salakoski**

University of Turku, Finland

## Abstract

Country images are increasingly popular but controversial policy concepts. The Western news media image of China is consequently a well-explored academic subject. This study contributes to the discussion by considering the news stream of an international news agency to contrast prior studies that analyze small-scale data sets. An automated dictionary method is proposed to analyze two Reuters data corpora RCV1 (1996–1998) and TRC2 (2008–2009) ( $N = 1,386,000$ ). The area under the receiver operating characteristic curve is employed and its development over time is statistically analyzed. China's media image was found to be relatively positive in comparison with Japan, South Korea, and Taiwan. The results also suggest that economy provides a greater positive impact on the China's image in Western media than culture, while politics has a negative impact. The results of the automated method are similar to those of a previous study in which a moderately sized data set was manually examined. This suggests that automated *sentiment* analysis can provide equally reliable observations as manual analysis but with smaller labor costs.

## Keywords

Automated content analysis, China, journalism, political communication

## Introduction

When conducting world politics, the image of a country in part determines the way in which that country is presented by the international media (Anholt, 2005). Therefore, concerns about one's own status are for many states as important as security and welfare goals (Gross-Stein, 2013). This makes states increasingly preoccupied by their images.

## Corresponding author:

Jukka Aukia, Centre for East Asian Studies, University of Turku, 20014 Turku, Finland.

Email: [jukka.aukia@utu.fi](mailto:jukka.aukia@utu.fi)

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>)

which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

China has, among others, emphasized traditional culture in its public diplomacy communication.<sup>1</sup> This is in line with the understanding that cultural communication as an element of public diplomacy attempts to convey a favorable image of national culture to facilitate diplomatic activities as a whole (Mitchell, 1986). While China has been polishing its international image since 1978, a turning point in the development of its international media relations took place between 2003 and 2004. At that time, according to some scholars, China's public diplomacy became increasingly sophisticated (J. Zhang, 2012, p. 303).

Although China's media image is a relatively well-explored academic subject, there are no empirical studies that use large-scale data sets to examine the changes in Western media coverage before and after policy change. Prior studies that specifically focus on China's media image mostly consist of small-to-moderate data sets. In these studies, the sample size ( $N$ ) ranges from 63 (Zhong & Zhang, 2016) to 4250 (Xiang, 2013). Moreover, the bulk of the prior studies have analyzed US-based specific newspapers such as the *New York Times* or the *Wall Street Journal*.

Rather than individual news media, we opted to analyze an international news agency (Reuters) in this study because it provides a contrast to the prior studies that analyzed the opinions of individual newspapers. It seems reasonable to assume that the content of individual newspapers is prone to be more emotional and opinion-based in comparison with international news agencies. Reuters' news articles are widely used in Western news outlets, and the analyzed articles are hence representative of a significant portion of the news streams which the general audience in Western countries is exposed to.<sup>2</sup>

We propose a method to analyze large-scale data sets because a large-scale automated analysis is preferable over a small-scale manual analysis due to the generally neutral tone of Reuters' news articles. Each article may contribute only slightly to the overall trend which is hence more reliably detected from a large sample. In addition, the large number of country–theme–time combinations that need to be considered in the analysis increases the sample size requirements. The large scale also allows to detect more subtle differences between countries and to investigate which news topics influence China's news media image.

In this article, we set out (1) to investigate what kind of impact, on an aggregate level, the image polishing efforts may have had on China coverage; (2) what specific news topics, including culture, influence the Western news media image of China; and (3) to compare the sentiment of China-related content to that of Japan, South Korea, and Taiwan.

To the best of our knowledge, no study to date has used automated statistical tools to analyze the development of country- and topic-specific *sentiment* over time in news coverage.<sup>3</sup> In this article, we apply an automated content analysis, a methodology which has gained attention within Political Science (Monroe & Schrodt, 2008). More specifically, the receiver operating characteristic (ROC) curve, calculated from the distributions of document sentiments, is applied as an appropriate measure and its development between corpora is statistically analyzed. The validity of the analysis is also evaluated.

We first review some general conceptual issues relating to state communication with foreign media and how this may be beneficial in achieving policy goals. After this, we look at the particular case of China. From these discussions we draw hypotheses to be tested with the proposed method. We then review prior methodological approaches, before finally presenting the actual analysis, results, and discussion.

### *Public diplomacy and mass media*

Before engaging in the policy choices China has made in its public diplomacy communication, we briefly review the general mechanisms state actors use to shape the policy environment of a foreign

country. We particularly highlight that prior research recognizes that mass media potentially influences both public opinion and policy elites. Both connections are mainly attributed to the fact that most people, whether decision-makers or not, lack direct experiences and hence are relying on the media to form opinions of other countries. The factual knowledge of public or individuals regarding foreign countries can indeed be limited, as, for instance, evidenced by the documented ignorance of the US public regarding foreign affairs (Carpini & Keeter, 1997).

Regarding the public, a seminal study by Perry found that in contrast to their personal knowledge, respondents relied on news to make inferences about foreign countries (Perry, 1985). More recent studies support the view (Entman, 2004; Wanta, Golan & Lee, 2004). Regarding agenda setting, prior research finds a correlation between the salience of foreign affairs in media and the salience of foreign affairs of the general public (Soroka, 2003; Wanta & Hu, 1993). Drawing from prior research, it is therefore possible to conceptually link media coverage and public opinion.

A number of studies show how the media also influences policy decision-makers. Studies that have interviewed policy-makers make the point that the news media is a major source of information for them (O’Heffernan, 1991). Considering the major developments in communication technologies including the Internet, studies show that both traditional and new media continue to shape the environment in which foreign policy is made (Potter, 2002). Like public, the agenda of policy elites can be influenced by making an issue salient (Liu, 2006). Media can also have an effect on how the policy-makers perceive an issue by applying different frames (Kiousis & Wu, 2008).

In sum, following Golan (2013), one can make the argument that nation-branding efforts do not seem to succeed if the media coverage and public perception of a country are negative. Political leaders also tend to pay attention to public attitudes when developing policies; thus, mass media and public opinion have the potential to influence foreign policy, especially in democratic countries (Holsti, 2004). These conceptual connections underline why states attach importance to their public diplomacy communication.

### *China’s relationship with Western media*

China’s public diplomacy efforts and the consequent relationship with Western media have received significant attention both inside and outside China. The research on the relations between PRC officials and Western media working in China highlights mutual mistrust and a controversial relationship. For many Chinese academics, the Western media is biased to the extent which implies a conscious effort to demonize China (see, for instance, Xie & Page, 2013). Chinese elites believe that global media stereotypes China as an “oriental” nation-state, offering a biased and negative perspective (Y. Wang, 2008). The globalization of state-run Chinese media corporations is also seen in China to face barriers from the global communications system that serves Western-based transnational monopolies. Thus China is not seen to be fully represented by global media due to the Western-dominated structure of information flow (Hu & Ji, 2012).

The controversial relationship is reflected in the opposing opinions held by Western and Chinese experts. For example, Zhang and Cameron refer to questionnaires conducted in bilateral communication conferences where American journalists and policy-makers viewed US coverage of China as balanced and objective, while most of their Chinese colleagues strongly disagreed (J. Zhang & Cameron, 2003, pp. 17–18). Moreover, several studies that address China’s cultural and political status insist that Western news media considers China to be culturally and politically inferior. For instance, a distorting and demonizing Western media bias is seen to stem from differences in news practice, ideology, and national interest (Xiang, 2013, p. 255). More specifically, the historical, cultural, and political “complexity” of China has been seen as posing challenges to especially US

journalists at whose mercy the national image of China is perceived to be (Peng, 2004, pp. 54–64). The controversy has also been seen to stem from “inherent tensions” that arise from the Chinese rhetoric as well as from the anti-communist ideology of the Western media (Zhong & Zhang, 2016).

Reflecting the attitudes described above, China under Hu Jintao devised the official policy doctrine of “Peaceful Rise” in 2003 to counter the Western “China threat theory.” This was continued in 2005 when the concept of “Harmonious World” was introduced as a proposal for a post-Western world order that would better accommodate different civilizations, including the non-liberal China. A related macro policy development took place in 2004 when China started developing the concept “soft power” as a policy tool to advance the understanding of China’s story. Soft power was raised to the fore of foreign policy in 2007.

Periodic policy efforts to appeal to foreign journalists intensified after the declaration of the “Peaceful Rise” in 2003. As part of the shift in the foreign policy, the foreign press core witnessed Chinese government public disclosure that was more transparent than before (D. Zhang, 2012, p. 687). In addition, China started a government spokesperson system that was fully implemented after the severe acute respiratory syndrome (SARS) epidemic in 2003. To many observers, it belongs among the central policies to relax the restrictions on foreign correspondents working in China (e.g. D. Zhang, 2012). Symbolizing the cultural “turn” in the Chinese public diplomacy, the Confucius Institute initiative has established Confucius Institutes around the world, the first of which was opened in South Korea in 2004 (Xie & Page, 2013, p. 856). In 2006, China further lifted restrictions on foreign correspondents working in China. The concessions were made permanent after the Beijing Olympics in 2008 (Sun, 2014, p. 3).

The PRC leadership has seemingly come to the conclusion that the strength of Chinese culture and its international influence are not congruent with China’s status (Hu & Ji, 2012). To support the view, there is evidence that Chinese culture tends to be admired by foreign publics, whereas its politics and governance are not (J. Wang, 2011, p. 6). For instance, it would seem that the perception of China along the cultural dimension is more favorable than along the political one in several Asian countries (Xie & Page, 2013, pp. 854–855). Moreover, Chinese soft power products tend to be viewed as propaganda in the West because they originate from a Communist regime (Chu, 2014).

China places its traditional culture into the core of its public diplomacy communication. Through this rehabilitation of traditional culture, “cultural soft power” has become popular in China’s articulations of itself, characterizing its public diplomacy communication (J. Wang, 2011). As a result, China has become a formidable cultural gatekeeper and producer that manages a vast publishing and broadcasting empire (Curtin, 2012). China’s leadership tends to assess that through boosting its international cultural influence it can achieve a competitive advantage and overcome the still largely negative image (Chu, 2014, p. 164).

To conclude, an argument can be made that China’s public diplomacy is media-centric with an emphasis on improving Western news media’s coverage of China and that the PRC public diplomacy communication centers on promoting China’s culture.

## **Hypotheses concerning China’s image in Western media**

Three hypotheses are proposed based on the above discussion. Note that only correlation is considered because the analysis of causality would require much more comprehensive data.

*Hypothesis 1 (H<sub>1</sub>).* PRC public diplomacy efforts regarding foreign journalists will be positively correlated with China coverage in terms of tone.

The tone of China-related coverage is expected to have been improved longitudinally if the policy approaches adopted by the PRC have been effective.

*Hypothesis 2 (H<sub>2</sub>).* Within China coverage, culture-related articles will have more positive tone than those related to politics.

It is assumed that the mostly liberal Western media does not react in a positive manner to authoritarian governance. Thus, coverage addressing Chinese culture is expected to be more positive in tone than coverage addressing Chinese politics. H<sub>2</sub> underlines the possibility that China's country image is divided.

*Hypothesis 3 (H<sub>3</sub>).* China-related coverage will be generally more negative in tone than coverage related to Japan, South Korea, or Taiwan.

Differences in political systems and orientations are assumed to influence Western media images: authoritarian China is expected to have more negative sentiment than the Western-style democracies of the comparison group.

## **Prior quantifications of China's image in Western media**

The quantification of China's media image has been the topic of several academic publications. The main body focuses on European and US media coverage of China. Most studies conclude that China is portrayed either selectively or negatively (see, for instance, Peng, 2004; Willnat & Luo, 2011; Xiang, 2013; Zhang & Cameron, 2003; Zhong & Zhang, 2016).

The most popular method to quantify the image of China in the international news media is manual content analysis. For instance, Zhang and Cameron (2003) analyze the content of three major US newspapers, the *Los Angeles Times*, the *New York Times*, and the *Washington Post* ( $N = 579$ ), while Golan and Lukito (2015) focus on the *New York Times* and the *Wall Street Journal* ( $N = 249$ ). The sizes of data sets range from small to moderate due to the fact that large data sets are intractable for manual analysis even if the text units are short. The range is well illustrated by the studies by Zhong and Zhang (2016) and Xiang (2013) which concern 63 and 4250 units of analysis, respectively.

The most relevant studies to this study are those that have a longitudinal approach. Peng (2004) observes that the coverage of China in the *New York Times* and the *Los Angeles Times* during the years 1992 and 2001 ( $N = 189$ ) increased, but the overall tone of reporting was negative. L. Zhang (2010) finds that (1) the overall proportion of Chinese coverage increased and (2) China drew the most attention in the economic and external political sectors in European transnational media (with the focus on the *Financial Times*, the *Economist*, and the *International Herald Tribune*) between the years 1989 and 2005 ( $N = 3004$ ). X. Wang and Shoemaker (2011) argue, based on the content analysis of the *New York Times*, the *Washington Post*, and *USA Today* between the years 1979 and 2007 ( $N = 1412$ ), that political freedom in China correlates with positive US media coverage which again correlates with favorable US public opinion on China. Finally, Yang and Liu (2012) observe that the "China threat" coverage in the *New York Times*, the *Washington Post*, the *Los Angeles Times*, the *Wall Street Journal*, and the *Chicago Tribune* between the years 1992 and 2006 ( $N = 376$ ) fluctuated

from category to category (China's escalating military power, China's increasing economic influence, and the political/ideological differences between China and the United States).

## Methods and data

The hypotheses of this study are examined with the comparisons of two statistical populations (e.g. China versus others) under specific circumstances. The unit of analysis is a news article. In the analysis, the sentiment score is a continuous interval variable, while the time point (before or after the policy changes), the topic (economy, politics, or culture), and the region (China, Japan, South Korea, or Taiwan)<sup>4</sup> of an article are nominal variables.

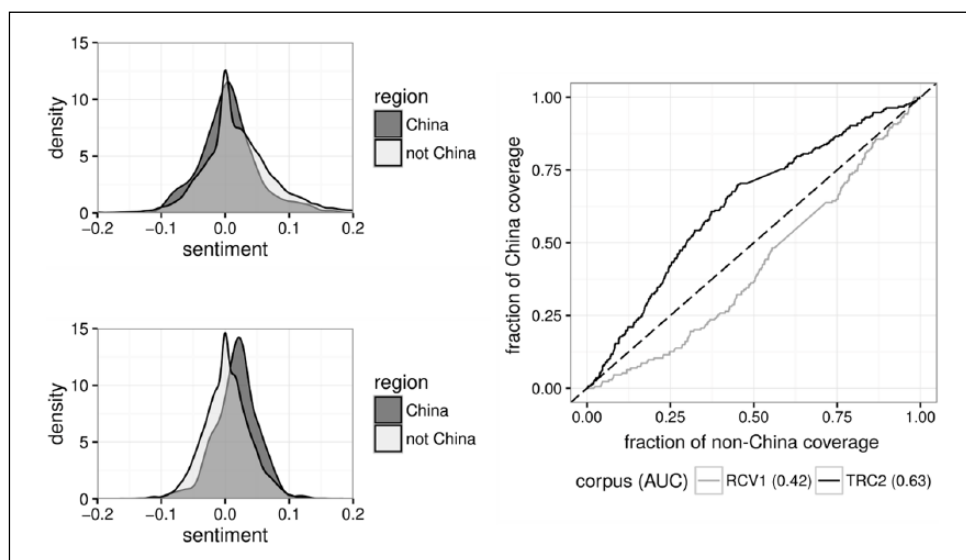
Years 2003 and 2004 are considered to be the key years in relaxing the restrictions on foreign journalists in China. Therefore, the Reuters Corpus, Volume 1 (RCV1) and Thomson Reuters Text Research Collection (TRC2) are appropriate as the two time points. They consist of news articles ranging from 20 August 1996 to 19 August 1997 and 1 January 2008 to 28 February 2009, respectively. These time periods are located before and after the beginning of the policy changes with a sufficient margin and should therefore reflect the development of the image of China due to the PRC public diplomacy efforts.

The dictionary approach suggested by Young and Soroka (2012) and discussed by Grimmer and Steward (2013) was adapted to compute the sentiment scores because such an approach is computationally inexpensive and can be fully automated with existing dictionaries. These two properties are prerequisites for large-scale analysis. The Harvard General Inquirer (GI) dictionary was used both for topic detection and sentiment analysis because the GI is "the oldest and most expansive dictionary" for political science and because it has both thematic and sentiment categories (Young & Soroka, 2012, p. 210).

The choice of the statistical method was influenced by the nature of text mining and the distribution characteristic that was considered to best reflect the aims of the analysis. Written communication involves complex phenomena. Consequently, distributions obtained from textual data are frequently non-parametric. This discourages from analyzing mean or median because a bias against a region may occur in some other form than a location shift. It would thus be a sub-optimal approach to use two-sample *t*-tests or linear models, for example. These methods also make assumptions on distributions which complicates their use in the case of non-parametric data, although some deviations from the assumptions may be acceptable in a large-scale analysis. The simultaneous analysis of multiple independent variables with linear models is useful when detailed information on their joint effects is needed. This study, however, focuses on the existence of a bias regardless of its form. As a result, an ROC curve analysis was chosen.

An ROC curve indicates how two distributions reside relative to each other. Each point on the curve corresponds to a value of a continuous variable, and the position of the point reveals the proportions of the two distributions located above the value. These proportions are always equal for two identical distributions, resulting in a diagonal ROC curve. If one distribution has some mass that is positively (or negatively) shifted, its proportion is higher (or lower), resulting in a curve residing above (or below) the diagonal. This is illustrated in Figure 1 in the case of sentiment distributions.

An ROC curve describes the separation of two continuous distributions. ROC curves are commonly used to determine how well predictive models can classify observations into two categories (indicated by a dichotomous variable) and to compare their goodness. The output of a model is a



**Figure 1.** An exaggerated example of sentiment distributions and ROC curves for illustrative purposes. *Upper left:* The sentiment distributions are close to each other in RCV1. *Lower left:* The sentiment distribution of China-related documents has shifted to the right (i.e. become more positive) relative to non-China-related documents in TRC2. *Right:* The more China coverage has been separated to the positive direction from the non-China coverage, the higher the ROC curve is located and hence the larger the area under the ROC curve (AUC). The dashed diagonal line indicates the case of no separation. The ROC curve of TRC2 is located higher than that of RCV1 because China coverage (relative to non-China coverage) is more positive in TRC2 than in RCV1.

continuous variable, and an ROC curve can be constructed from the two output distributions conditioned on the dichotomous variable. The more the conditional distributions are separated, the better discrimination performance the model has. Note that the ROC analysis concerns the prediction of a dichotomous value using the continuous output of a model. This is opposite to linear models which model a continuous dependent variable in terms of (possibly) dichotomous independent variables.

In this study, the predictive model (i.e. the sentiment score) is predetermined and its discrimination performance is compared across conditions (i.e. the selected subsets of data). The observed differences of ROC curves hence reflect the differences in data: if the sentiment score can predict whether a news article is about a particular region or not, there is some bias for or against that region. The advantage of the ROC analysis is that an ROC curve addresses all kinds of differences (and not only, for example, mean or median). The analysis can also avoid making assumptions regarding the distributions.

The separation of two distributions can be summarized by the area under the ROC curve (AUC) as elaborated in Figure 1. Identical distributions yield  $AUC = 0.5$ , and the extreme cases  $AUC = 0$  and  $AUC = 1$  indicate complete separation. This study uses the empirical AUC which is appropriate for non-parametric distributions and equivalent to the well-known Wilcoxon–Mann–Whitney  $U$ -statistic that was first proposed by Wilcoxon in 1945.

In this study, an AUC value calculated from sentiment scores measures the separation of a given region under specific temporal and thematic conditions (target distribution) from the other regions under the same conditions (reference distribution). One region is hence essentially evaluated against the other regions. For example, the AUC value calculated from China-related documents in RCV1 (target distribution) and non-China-related documents in RCV1 (reference distribution) reveals whether China coverage had the tendency of being more negative (or more positive) than coverage in general in 1997. AUC values are consequently baseline-adjusted sentiment scores that isolate the effect of the region from the other effects.

Regional, thematic, and temporal differences may be analyzed by comparing AUC values using the statistical test developed by DeLong, DeLong, and Clarke-Pearson (1988). The test examines whether two AUC values can be considered unequal. For example, the difference between China's culture AUC in 1997 and 2008 shows the development of China's cultural image over time. Note that if the China-related culture documents in RCV1 and in TRC2 were directly compared, the observed differences would also include the effect of the development of news reporting in general (such as journalistic guidelines and style). The DeLong test was originally developed for paired observations but later extended to unpaired observations in the pROC package (Robin et al., 2011) for the R programming language.

The analysis relies on AUC values that are validated to reflect the specific regions and conditions as well as the image in Western media. The categorization was evaluated by the hand coding of a random sample by domain experts. The concurrent validity of the sentiment AUC as a measure of media image was evaluated by calculating the correlation of the sentiment ranking of countries to the Transparency International Corruption Perceptions Index and to the World Bank gross domestic product (GDP) per capita rankings.<sup>5</sup> Prior studies suggest that international news coverage associates positively with such country characteristics as economic development (see, for instance, X. Wang & Shoemaker, 2011, p. 3). There is also a negative correlation between the level of corruption and GDP per capita (Mauro, 1995; 2004). Therefore, sentiment AUC is expected to correlate with corruption and GDP per capita rankings if it reflects media image.

The following sections describe the details of the four steps of the analysis.

### *Data preparation*

The documents were cleaned from inline metadata and tokenized. The words were then lemmatized and converted to lower case.<sup>6</sup> The total frequency of words and the frequencies of the words belonging to the selected categories of the GI were recorded for each document.

### *Document categorization and filtering*

Each document was assigned to exactly one region and zero or more topics. While RCV1 includes the topic and region categorizations in its metadata, TRC2 does not have such information. To produce comparable results, the region information was extracted from TRC2 documents by automatically detecting the city–date pattern in the dateline, such as “TOKYO, Jan 1 (Reuters),” and determining the region from the city. The corresponding information in RCV1 is readily available as the location of the document creator as a piece of metadata. For both corpora, a document was assigned to a particular topic if its headline contained at least one word belonging to the corresponding GI category. The Econ\*, Polit\*, and Exprs categories were chosen to represent the



economy, politics, and culture topics, respectively. This procedure is likely to produce a categorization that is less accurate than the metadata of RCV1 but obtainable for both corpora.

The data sets were filtered to improve the reliability of the analysis. First, the documents without a region or without full sentences were excluded. Such documents are, for example, service alerts and commodity price updates. Second, the regions that contributed less than 0.1% to the number of documents were removed because an AUC value calculated from a small number of documents would contain unacceptable amount of noise that would disturb the analysis.

### Sentiment analysis

The document sentiment score  $q$  measures the tone of a document and, following Young and Soroka (2012), is defined as the difference between positive and negative words normalized by the length of the text. It is formally written as

$$q(d) = \frac{w_d^+ - w_d^-}{w_d}$$

where the frequencies of the positive, negative, and all words of the document  $d$  are denoted by  $w_d^+$ ,  $w_d^-$ , and  $w_d$ , respectively. The words with positive and negative connotations were chosen to be represented by the GI categories Pos and Neg, respectively.

The AUC values were first computed for each combination of region, topic, and time point. The DeLong test was then applied to the pairs relevant for the three hypotheses. The two time periods of each region were compared under each topic and overall ( $H_1$ ), China's culture topic was compared to China's economy and politics topics in both time periods ( $H_2$ ), and China was compared to Japan, South Korea, and Taiwan in both time periods under each topic and overall ( $H_3$ ).

### Validation

A stratified random sample of documents was selected for hand coding from the computer-categorized data set such that there were 15 documents for each of the 30 possible corpus–topic–region combinations. Three domain experts examined each document and assessed whether the computer-assigned topic and region were correct. The precision of the categorization was estimated from the consensus as the proportion of correctly assigned documents. The quality of the hand coding was evaluated with Fleiss'  $\kappa$  (Fleiss, 1971) which quantifies an inter-coder agreement between more than two coders and takes into account the agreement by chance.

Spearman's rank correlation (Spearman, 1904) of sentiment AUC to Transparency International Corruption Perceptions Index and to World Bank GDP per capita rankings was calculated. The data from years 1997 and 2008 were used to match the time periods of RCV1 and TRC2, respectively.

## Results

The automated content analysis provides insights into the differences in tone across regions and topics as well as into the development over time. Below, the validity of the results is first established and the analysis of AUC values is then presented.

**Table 1.** Precision of automated categorization as evaluated by the majority vote of three domain experts.

	Economy	Politics	Culture	Total
China	76.7% (0.72)	50.0% (0.64)	36.7% (0.70)	91.1% (0.51)
Japan	80.0% (1.00)	40.0% (0.72)	3.3% (0.15)	80.0% (0.70)
South Korea	90.0% (0.52)	60.0% (0.82)	23.3% (0.94)	93.3% (0.45)
Taiwan	90.0% (0.71)	46.7% (0.60)	13.3% (0.66)	90.0% (0.75)
Total	89.3% (0.87)	46.7% (0.70)	21.3% (0.70)	

The values in parenthesis are Fleiss'  $\kappa$  scores.

### Validation

The results of the hand coding (Table 1) show that all regions were categorized with high precision (80%–93%). This validates the assumption that the region from which the report originates reflects the perspective of the article, and the sentiment should hence be associated with that region instead of any other or all the locations mentioned in the article. The observed precision for the economy topic is good (89%), but those for the politics (47%) and culture topics (21% overall but 37% for China) indicate that the documents of these topics are only moderately enriched in the corresponding subsets. Particular criticism should therefore be exercised when analyzing these subsets, although low precision tends to decrease the sensitivity of the analysis to detect differences instead of invalidating the whole analysis (see section “Discussion”).

Fleiss'  $\kappa$  values in Table 1 indicate that with the exception of the culture of Japan, the hand coding can be considered sufficiently reliable. According to Landis and Koch (1977), Fleiss'  $\kappa$  values above 0.60 already indicate substantial agreement. However, since  $\kappa$  values are affected by skewed distributions of categories (Di Eugenio & Glass, 2004), these thresholds must be applied with caution. Indeed, the proportion of cases in which all three annotators agree is generally above 80% (data not shown), including those subsets that have only moderate  $\kappa$  values.

The country-specific tone of Reuters reporting correlates with national corruption and national GDP per capita levels, which implies that the sentiment AUC is an appropriate quantity for the analysis. The observed correlation of the Transparency International Corruption Perceptions Index with AUC was 0.55 for RCV1 and 0.41 for TRC2, while that of the World Bank GDP per capita ranking was 0.54 for RCV1 and 0.54 for TRC2. These correlations were statistically significant ( $p < 0.01$ ).

### Volume of coverage

The overall relative document counts (Table 2) show that the volume of China coverage doubled over time, while of the three comparison regions only South Korea exhibits an increase of similar magnitude. The topic-specific relative document counts (Table 4) further indicate that the increase was notably greater for the culture topic (+170%) than for the economy and politics topics (+87% and +99%, respectively).

### Hypotheses

To get a sense of the AUC magnitudes in Tables 2 to 4, as an illustrative example, consider that Belgium (a first-world country with high GDP and low corruption) has AUC 0.569 in RCV1 and

**Table 2.** Results of the overall analysis.

	Document count		AUC		Temporal p-value		Regional p-value	
	RCVI	TRC2	RCVI	TRC2			RCVI	TRC2
China	11,439 (1.54%)	20,659 (3.21%)	0.481 (53.7%)	0.508 (62.1%)	$4.20 \times 10^{-17}$		—	—
Japan	41,259 (5.56%)	24,124 (3.75%)	0.482 (54.9%)	0.455 (32.8%)	$2.83 \times 10^{-30}$		$6.41 \times 10^{-1}$	$9.22 \times 10^{-87}$
South Korea	6155 (0.83%)	11,498 (1.79%)	0.423 (20.7%)	0.428 (25.9%)	$2.40 \times 10^{-1}$		$2.48 \times 10^{-37}$	$3.96 \times 10^{-141}$
Taiwan	4468 (0.60%)	5661 (0.88%)	0.410 (14.6%)	0.502 (55.2%)	$1.36 \times 10^{-73}$		$1.62 \times 10^{-52}$	$1.40 \times 10^{-1}$
Total	742,323	643,715						

For each region, the absolute and relative document counts as well as the area under the receiver operating characteristic curve (AUC) values and their relative ranks (i.e. the proportion of regions with equal or smaller AUC) are reported for both corpora. The p-values refer to the DeLong tests: a temporal p-value concerns the difference between corpora within a region, while a regional p-value concerns the difference between China and another region within a corpus.

**Table 3.** Results of the analysis of China’s topics.

	AUC		<i>p</i> -value	
	RCV1	TRC2	RCV1	TRC2
Culture	0.459	0.520	–	–
Economy	0.451	0.508	$5.54 \times 10^{-1}$	$1.37 \times 10^{-1}$
Politics	0.490	0.456	$3.48 \times 10^{-2}$	$2.80 \times 10^{-11}$

The *p*-values refer to the DeLong tests and concern the differences between China’s culture topic and its other topics. The area under the receiver operating characteristic curve (AUC) values are repeated from Table 3 for clarity.

AUC 0.661 in TRC2, while the corresponding values for Kenya (a third-world country with low GDP and high corruption) are 0.334 and 0.399.

The overall tone of China coverage improved between the analyzed time periods, which can be seen both in the *p*-value ( $p < 0.0001$ ) and in the relative AUC ranks (a change from 54% to 62%) in Table 2. This *observation supports the hypothesis H<sub>1</sub>*. The *p*-values in Table 3 indicate that the culture topic has lower AUC in RCV1 ( $p = 0.035$ ) but higher AUC in TRC2 ( $p < 0.0001$ ) than the politics topic, albeit the former *p*-value is not significant if the conservative Bonferroni correction ( $N = 44$ ) is applied. These *results support the hypothesis H<sub>2</sub>* that the coverage of culture was more positive in tone than the coverage of politics. A difference was not observed between economy and culture. China’s AUC values are among the highest in Table 4. In all pairwise comparisons, either China’s AUC is the higher one with a statistically significant *p*-value or there is no statistical evidence for difference. Therefore, *the data do not support the hypothesis H<sub>3</sub>* that China-related coverage was more negative in tone than coverage related to the comparison regions. In addition, the relative AUC ranks (54% and 62%, respectively) indicate that the tone of China coverage was above the median.

The observed change in AUC is even greater for China’s economy ( $p < 0.0001$ ) and culture ( $p < 0.0001$ ) than overall. In contrast, the change in China’s politics AUC is in the opposite direction ( $p = 0.00023$ ). Of the comparison group, only the positive change in overall AUC of Taiwan ( $p < 0.0001$ ) is comparable to that of China. In contrast to China, the change in Taiwan’s AUC is positive both in the economy category ( $p < 0.0001$ ) and in the politics category ( $p < 0.0001$ ).

**Discussion**

This study set out with the aim of assessing the image of China in the Western news media. The literature review revealed that the existing body of research consists of hand-coded small-scale studies and that very few studies concern methods that would enable large-scale inquiries. Therefore, this study was designed (1) to determine the tone of overall China-related coverage and (2) to detect specific topics and determine their tone in China coverage in a large data set. Regarding the PRC policy changes and the increase in PRC public diplomacy activities, this study focused on correlation with sentiment rather than causality.

*Validity*

The observation that the AUC ranking correlates both with the Transparency International Corruption Perceptions Index and with the World Bank GDP per capita ranking suggests that AUC

**Table 4.** The results of the topic-specific analyses.

	Document count		AUC		Temporal p-value	Regional p-value	
	RCV1	TRC2	RCV1	TRC2		RCV1	TRC2
<i>Economy</i>							
China	6027 (1.72%)	11,089 (3.21%)	0.451 (31.7%)	0.508 (55.6%)	1.27 × 10 <sup>-36</sup>	—	—
Japan	12,436 (3.56%)	12,587 (3.65%)	0.461 (35.4%)	0.453 (29.6%)	2.88 × 10 <sup>-2</sup>	2.49 × 10 <sup>-2</sup>	6.87 × 10 <sup>-51</sup>
South Korea	3248 (0.93%)	6367 (1.85%)	0.425 (18.3%)	0.450 (25.9%)	4.55 × 10 <sup>-5</sup>	2.06 × 10 <sup>-5</sup>	3.58 × 10 <sup>-42</sup>
Taiwan	2472 (0.71%)	2922 (0.85%)	0.395 (11.0%)	0.498 (46.3%)	1.29 × 10 <sup>-46</sup>	1.30 × 10 <sup>-18</sup>	8.23 × 10 <sup>-2</sup>
Total	349,761	344,993					
<i>Politics</i>							
China	1564 (1.41%)	2362 (2.80%)	0.490 (52.2%)	0.456 (33.3%)	2.25 × 10 <sup>-4</sup>	—	—
Japan	2220 (2.01%)	1939 (2.30%)	0.476 (50.0%)	0.455 (31.9%)	1.52 × 10 <sup>-2</sup>	1.38 × 10 <sup>-1</sup>	8.91 × 10 <sup>-1</sup>
South Korea	965 (0.87%)	1086 (1.29%)	0.362 (14.4%)	0.413 (27.8%)	2.13 × 10 <sup>-5</sup>	8.69 × 10 <sup>-30</sup>	1.98 × 10 <sup>-5</sup>
Taiwan	558 (0.50%)	413 (0.49%)	0.424 (23.3%)	0.527 (73.6%)	1.35 × 10 <sup>-10</sup>	1.53 × 10 <sup>-7</sup>	1.68 × 10 <sup>-7</sup>
Total	110,554	84,221					
<i>Culture</i>							
China	506 (1.62%)	1378 (4.38%)	0.459 (39.5%)	0.520 (67.3%)	4.25 × 10 <sup>-5</sup>	—	—
Japan	1115 (3.56%)	1115 (3.54%)	0.556 (86.4%)	0.441 (32.7%)	7.06 × 10 <sup>-21</sup>	2.94 × 10 <sup>-10</sup>	7.45 × 10 <sup>-12</sup>
South Korea	336 (1.07%)	448 (1.42%)	0.588 (95.1%)	0.389 (12.7%)	7.84 × 10 <sup>-18</sup>	2.53 × 10 <sup>-8</sup>	8.44 × 10 <sup>-19</sup>
Taiwan	159 (0.51%)	257 (0.82%)	0.409 (24.7%)	0.456 (40.0%)	8.23 × 10 <sup>-2</sup>	4.36 × 10 <sup>-2</sup>	7.30 × 10 <sup>-4</sup>
Total	31,292	31,459					

For each region, the absolute and relative document counts as well as the area under the receiver operating characteristic curve (AUC) values and their relative ranks (i.e. the proportion of regions with equal or smaller AUC) are reported for both corpora. The p-values refer to the Delong tests: a temporal p-value concerns the difference between corpora within a region, while a regional p-value concerns the difference between China and another region within a corpus.

quantifies the tone of document sets. The abundance of economy-themed articles in Reuters' coverage may be a significant factor in this correlation. However, the conclusions regarding China's culture coverage should be treated with caution due to low categorization precision. The low precision leads to a distribution that resembles the overall distribution rather than the distribution of culture documents. Given the dominance of economy coverage in the data, such a distribution is a mixture of the economy and culture topics, and the observed AUC subsequently reflects this mixture. The hypothesis  $H_2$  concerning the role of culture in the overall media image should hence be further investigated. In particular, various types of textual data (such as news articles and social media posts) should be examined to discover whether they differ in the tone or prevalence of the economy, politics, and culture topics.

### *Hypotheses*

The results reveal a positive development of tone in China-related overall coverage between 1996 and 2008. China's culture and economy were equally favorably received in Western media and had similar positive development, although the conclusions regarding culture need further validation. The positive development of economy is likely related to the increase in China's economic weight between 1996 and 2008. As a result of their negative development between the corpora, the politics-themed articles were found to be more negative in tone than the economy- or culture-themed articles in 2008. This study therefore agrees with the study by Xiang (2013) in that China-related economic and cultural topics are more favorably presented than political ones. The conducted experiment shows that, in comparison with Japan, South Korea, and Taiwan, China's media image is relatively positive.

The changes in the volume of China coverage indicate a growth of interest in China in general and in Chinese culture in particular. The observed abundance of economy coverage over culture coverage in the Reuters data, however, suggests that culture is a marginal topic in global news coverage. This supports the meta-study of Wu (2000), which showed both economic issues and political affairs dominating international news agenda. As a result, it may be that in addition to the negative tone of China-related political news, it is the large volume of the economy-related articles that drives the development of China's Western media image. It is therefore interesting to note Brady's (2006) observation, according to which the guidelines of the Central Propaganda Department overseeing PRC public diplomacy include also the talking up of economy. Also, according to Ferdinand (2016), economic reformers within the PRC have seemed to envision market reforms as a source of China's international attractiveness. In any case, the present findings are consistent with those of Ahern (1984) and Kim and Barnett (1996) who found that the amount of economic clout is positively related to the amount of coverage a country receives in foreign news media.

### *Methodology*

The ROC curve analysis has two major advantages. First, AUC is an applicable quantity regardless of the distribution of the analyzed quantity. This is useful in analyzing complex phenomena in textual data which commonly involve non-parametric distributions. Second, AUC isolates the effect of interest because an observed AUC value reflects the property that distinguishes the target and reference distributions. The DeLong test can be applied after this baseline adjustment to evaluate the differences between two target distributions.

A disadvantage of AUC is that the observed values are non-trivial to interpret. The AUC values do not have an intuitive meaning that directly describes the text (cf. the sentiment score  $q$ ), but rather AUC describes a statistical property of the quantity. The observed range of AUC values may vary across data sets and analyses, which makes the evaluation of the practical significance of observed differences challenging. To alleviate this issue, the relative ranks of regions (among all regions under the same conditions) are also reported in this study.

Due to its comparative nature, the analysis has resistance against potential biases caused by imperfect categorization. Biases are accounted for in the comparison of the ROC curves provided that they are similar across the subsets. This can occur when the categorization precisions are similar and errors random. However, the sensitivity of the analysis diminishes as the categorization precision decreases because the reference and target distributions become partially mixed. The evaluation of the automatic categorization should be extended in future studies to quantify these effects.

### *Implications*

The examined large-scale data support the view that China-related coverage is neutral or positive in tone which, at first, seems to contradict the prior studies that have found negative tones in small samples and neutral tones in moderately sized samples analyzing individual news sources such as the *New York Times* or the *Economist*. However, we maintain that results from samples that greatly differ in size and in the type of source material should not be directly compared. A small-scale study using a manual analysis can discover negative tone in some specific forum or topic, while a statistical large-scale study, such as this, can still be correct in maintaining that the overall tone is positive.

### *Future directions*

The proposed application of the ROC analysis could be used in a variety of research designs in political studies that use text as primary data. It could be particularly beneficial in testing hypotheses that address the longitudinal development or cross-sectional comparison of categorized content that involves non-parametric distributions and requires baseline adjustments. To give an example, the ROC analysis could be used to test to which direction the tone of US senatorial speeches by republican representatives developed between the Bush and the Obama administrations in comparison with the speeches held by their democratic colleagues. In this case, it would be problematic to conceive an appropriate measurement baseline for the development of tone, but the ROC analysis could still be applied.

Improvements in the sentiment analysis and categorization quality could be achieved through dictionary refinement efforts (e.g. Young & Soroka, 2012), machine learning methods, or the use of syntactic analyses (e.g. Di Caro & Grella, 2013) to generalize the method. The validation through correlation with the corruption and GDP rankings proposed in this study may also indicate an interesting prospect for concurrent statistical validation.

### **Conclusion**

This study contributes to the literature on China's Western news media image by proposing an alternative approach to this relatively well-explored subject. While the news media image of China

was found to be more positive than expected, the results also suggest that instead of the culture-themed articles, it may be that the volume of the economy-themed articles and the negative tone of the politics-themed articles are the major contributors to the news media image of China. In contrast to *manual sentiment analysis*, nonetheless, the automated method proposed here is less labor-demanding but requires advanced validation measures.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Notes

1. For the role of culture in China's public diplomacy, see Hu and Ji (2012).
2. Reuters is one of the three major Western news agencies that not only produces and gathers reporting but also sells them to subscribing news organizations. Others being Associated Press (USA) and Agence France-Presse (France). Reuters News (headquarters in London, England) is a subsidiary of the Thomson Reuters Corporation (headquarters in New York, USA, and Toronto, Canada), making the agency transnational surpassing individual countries and individual newspapers.
3. The methodological contribution of this study concerns automated analysis of textual sentiment. Regarding topic detection and categorization, existing studies have applied more sophisticated methods (see, for instance, Lucas et al. 2015; Quin, Monroe, Colaresi, Crespin, & Radev, 2010).
4. In the preliminary experiments, Hong Kong was also considered as a region, but it was dropped from the analysis after observing that it could not be reliably categorized by the automated method. This is not very surprising since Hong Kong has been the regional base for Western outlets to cover Asia Pacific, East, and South East Asia.
5. Transparency International Corruption Perceptions Index. 2014. Retrieved from <http://www.transparency.org/research/cpi/overview>. The World Bank GDP per capita rankings. Retrieved from <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>.
6. Punctuation tokens (i.e. tokens without a letter or a digit) were filtered out along with common stop words.

## References

- Ahern Jr. T. (1984). Determinants of foreign coverage in newspapers. In R. L. Stevenson, & D. L. Shaw (Eds.), *Foreign news and the new world information order* (pp. 217-236). Ames: Iowa State University Press.
- Anholt, S. (2005). Nation brand as context and reputation. *Place Branding and Public Diplomacy*, 1, 224–228.
- Brady, A.-M. (2006). Guiding hand: The role of the CCP central propaganda department in the current era. *Westminster Papers in Communication and Culture*, 3(1), 58–77.
- Carpini, M., & Keeter, S. (1997). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.
- Chu, Y. (2014). The politics of reception: “Made in China” and Western critique. *International Journal of Cultural Studies*, 17, 159–173.
- Curtin, M. (2012). Symposium: Chinese media and globalization. *Chinese Journal of Communication*, 5(1), 1–9.
- DeLong, E., DeLong, D., & Clarke-Pearson, D. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837–845.
- Di Caro, L., & Grella, M. (2013). Sentiment analysis via dependency parsing. *Computer Standards & Interfaces*, 35, 442–453.
- Di Eugenio, B., & Glass, M. (2004). The Kappa statistic: A second look. *Computational Linguistics*, 30, 95–101.
- Entman, R. (2004). *Projections of power: Framing news, public opinion, and U.S. Foreign policy*. Chicago, IL: University of Chicago Press.



- Ferdinand, P. (2016). Westward ho—The China dream and “one belt, one road”: Chinese foreign policy under Xi Jinping. *International Affairs*, 92, 941–957.
- Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 378–382.
- Golan, G. (2013). Introduction: An integrated approach to public diplomacy. *American Behavioral Scientist*, 57, 1251–1255.
- Golan, G., & Lukito, J. (2015). The rise of the dragon? Framing China’s global leadership in elite American newspapers. *The International Communication Gazette*, 77, 753–771.
- Grimmer, J., & Steward, B. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21, 267–297.
- Gross-Stein, J. (2013). Threat perception in international relations. In L. Huddy, & D. Sears (Eds.), *The Oxford handbook of political psychology* (pp. 364–394.). New York, NY: Oxford University Press.
- Holsti, O. (2004). *Public opinion and American Foreign policy*. Ann Arbor: University of Michigan Press.
- Hu, Z., & Ji, D. (2012). Ambiguities in communicating with the world: The “going-out” policy of China’s media and its multilayered contexts. *Chinese Journal of Communication*, 5, 32–37.
- Kim, K., & Barnett, G. (1996). The determinants of international news flow: A network analysis. *Communication Research*, 23, 323–352.
- Kiousis, S., & Wu, X. (2008). International agenda-building and agenda-setting: Exploring the influence of public relations counsel on US news media and public perceptions of Foreign nations. *International Communication Gazette*, 70, 58–75.
- Landis, R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Liu, X. (2006). *Modeling bilateral international relations: The case of US-China interactions*. New York, NY: Palgrave Macmillan.
- Lucas, C., Nielsen, R., Roberts, M., Stewart, B., Storer, A., & Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis*, 23, 254–277.
- Mauro, P. (1995). Corruption and growth. *Quarterly Journal of Economics*, 110, 681–712.
- Mauro, P. (2004). The persistence of corruption and slow economic growth. *IMF Staff Papers*, 51(1), 1–18.
- Mitchell, J. (1986). *International cultural relations*. London, England: Allen & Unwin.
- Monroe, B., & Schrod, P. (2008). Introduction to the special issue: The statistical analysis of political text. *Political Analysis*, 16, 351–355.
- O’Heffernan, P. (1991). *Mass media and American Foreign Policy: Insider perspectives on global journalism and the Foreign policy process*. Norwood, OH: Ablex.
- Peng, Z. (2004). Representation of China: An across time analysis of coverage in the New York Times and Los Angeles Times. *Asian Journal of Communication*, 14, 53–67.
- Perry, D. (1985). The mass media and inference about other nations. *Communication Research*, 12, 595–614.
- Potter, E. (Ed.). (2002). *Cyber-diplomacy: Managing Foreign policy in the twenty-first century*. London, England: McGill-Queen’s University Press.
- Quin, K., Monroe, B., Colaresi, M., Crespin, M., & Radev, D. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54, 209–228.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J-C., & Muller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77.
- Soroka, S. (2003). Media, public opinion and foreign policy. *The Harvard International Journal of Press/Politics*, 8, 27–48.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.
- Sun, W. (2014). Configuring the foreign correspondent: New questions about China’s public diplomacy. *Place Branding and Public Diplomacy*, 8(October), 1–14.
- Wang, J. (2011). Introduction: China’s search of soft power. In J. Wang (Ed.), *Soft power in China: Public diplomacy through communication*. (pp. 1–18.). New York, NY: Palgrave Macmillan.

- Wang, X., & Shoemaker, P. J. (2011). What shapes American's opinion of China? Country characteristics, public relations and mass media. *Chinese Journal of Communication*, 4(1), 1–20.
- Wang, Y. (2008). Public diplomacy and the rise of Chinese soft power. *Annals of the American Academy of Political and Social Science*, 616, 257–272.
- Wanta, W., Golan, G., & Lee, C. (2004). Agenda setting and international news: Media influence on public perceptions of foreign nations. *Journalism & Mass Communication Quarterly*, 81, 364–377.
- Wanta, W., & Hu, Y.-W. (1993). The agenda-setting effect of international news coverage: An examination of different news frames. *International Journal of Public Opinion Research*, 5, 250–264.
- Willnat, L., & Luo, Y. (2011). Watching the dragon: Global television news about China. *Chinese journal of communication*, 4, 255–273.
- Wu, H. (2000). Systemic determinants of international news coverage: A comparison of 38 countries. *Journal of Communication*, 50, 110–130.
- Xiang, D. (2013). China's image on international English language social media. *The Journal of International Communication*, 19, 252–271.
- Xie, T., & Page, B. (2013). What affects China's national image? A cross-national study of public opinion. *Journal of Contemporary China*, 22, 850–867.
- Yang, Y. E., & Liu, X. (2012). The "China Threat" through the Lens of US print media: 1992–2006. *Journal of Contemporary China*, 21, 695–711.
- Young, L., & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, 29, 205–231.
- Zhang, D. (2012). A relational perspective on media relations strategies: The Chinese government's news conferences from 2001 to 2009. *Public Relations Review*, 38, 684–696.
- Zhang, J. (2012). Making sense of the changes in China's public diplomacy: Direction of information flow and messages. *Place Branding and Public Diplomacy*, 4, 303–316.
- Zhang, J., & Cameron, G. (2003). China's agenda building and image polishing in the US: Assessing an international public relations campaign. *Public Relations Review*, 29, 13–28.
- Zhang, L. (2010). The rise of China: Media perception and implications for international politics. *Journal of Contemporary China*, 19, 233–254.
- Zhong, L., & Zhang, J. (2016). Political myth as strategic communication: Analysis of Chinese dream's rhetoric and English news media's interpretation. *International Journal of Strategic Communication*, 10, 51–68.

### Author biographies

Jukka Aukia is a postdoc researcher at the Centre for East Asian Studies at the University of Turku. His main research interests include China's soft power, media, and public diplomacy.

Juho Heimonen is a PhD student in Computer Science at the University of Turku. His main research interests are ontologies, machine learning, and text mining in the bio-health domain.

Tapio Pahikkala is an assistant professor of Computer Science at the University of Turku. His work focuses on the development and applications of machine learning techniques.

Tapio Salakoski is a professor of Computer Science at the University of Turku. His research interests include text mining in the biomedical domain and the possibilities of information and language technologies in improving the quality of health care.