Modeling the Global News Flows:

A time-lag analysis of international news flows with a news categorization program

Kohei Watanabe

September, 2012

Using a newly developed news categorization algorithm and cross-correlation analysis, more than two hundred thousand international news items, which were collected from eight countries across the world, were analyzed in terms of global media agenda-setting. The analysis achieved constructing a model of international news circulation by estimating time-lags with which international news flows between different types of news providers. The model revealed the order and speed of global news flows from news agencies to elite newspapers, national newspapers and news portal sites, showing important roles of both news agencies and elite newspapers for national newspapers in deciding coverage of minor countries, which rarely become origins of international news, and progressively intensifying concentration of news coverage to major countries as news flows down to the lower tiers.
In 1992, Livingston has stated “international agenda-setting remains one of the least studied and least understood processes of international politics” (Livingston, 1992, p. 313). It was two decades ago when he made the statement, but the situation does not seem to have changed much. Despite its political importance, especially when we are constantly facing social, economic and ecological issues, such as human rights violation, world financial crisis and climate change, which demand international cooperation to tackle, the global agenda-setting process was not been studied much, because, as Groshek (2008) points out, the study of the process is a difficult task given the speed of information flows on electronic media and scale of the international news collection and distribution system. However, thanks to the Internet, availability of news items in digital formats had dramatically increased and costs to collect news from across the world had sharply dropped, so a large-scale research on global news flows has become possible. In this article, a new method of a large scale analysis of international news, which fully capitalizes on the recent development in computer text-analysis techniques in political science, will be presented and international news flows will analyzed based on the classic framework of agenda-setting research developed by McCombs and Shaw.

Literature Review

Since the seminal research by McCombs and Shaw (1972) about the influence of TV and newspapers on the voters' perception of important political issues during the presidential campaign in 1968, while many agenda-setting research projects were carried out on domestic news, some researchers were paying attention to agenda-setting effects of foreign news. Page and Shapiro (1992) argue that media effects in shaping public opinion on foreign issues become stronger than domestic ones because media is the only source of knowledge about foreign events, and found changes in public opinion among the US audience as a result of news coverage on international issues. Iyengar and Simon (1993) have found strong correlation between the amount of news coverage of the Gulf War on ABC News and the number of respondents in public opinion polls who had answered that the Gulf War is the most important issue; the heavy coverage on the war had raised the salience of the issues, while it had overshadowed other issues regarding the state of US economy, drugs and budget deficit, which had been salient before the war had started. More recently, Wanta, Golan, and Lee (2004) reported time-lag correlations between numbers of mentions of countries on TV news programs and people's perception of the importance of the countries.

Changes in public opinion by media coverage and their political implications in
foreign policy making have been discussed extensively under the ‘CNN effect’ thesis since the 1990s. The proponents of the effect have argued that extensive media coverage of conflicts had led the US government to military’s interventions in Haiti, Somalia, Bosnia and Kosovo, and its withdrawal from Somalia, in the 1990s (Bahador, 2007; Belknap, 2001; e.g. Robinson, 2002), but Gilaboa (2005), after reviewing studies on the effect, pointed out the lack of scientific evidence of such an effect. Nevertheless, the lack of strong evidence does not mean that the such effect does not exist, because the lack of strong evidence might be due to the small number of the cases of military interventions: most recently, researchers found correlations between the level of coverage of foreign disasters on the US media and the amount of aid from the US government for the affected countries by using large-n statistical analyses (Eisensee & Strömberg, 2007; Van Belle, 2008).

Emergence of the ‘CNN effect’ thesis is due to the fact that the development of the satellite communication system had enabled faster dissemination of news from anywhere. Before the era of satellite communications, the world-wide circulation of news had been totally in the hands of international news agencies, namely Reuters, AP, and AFP. Galtung and Ruge (1965) found that 87% of the articles about the Congo, Cuban and Cyprus crises in the 1960s on Norwegian newspapers were provided by AP, UPI, Reuter, and AFP. Salamore (in Boyd-Barrett, 1980) conducted content analysis of the top newspapers in India, Kenya, Lebanon, Japan, and Norway every month between 1961 and 1968, and more than half of the international news items were found to be provided by the four major news agencies. Matta’s content analysis of 16 daily newspapers in 14 Latin American countries in 1975 showed that 80% of the foreign news items were from the Western news agencies (in UNESCO, 1985). Schramm (1978) performed content analysis of articles about non-Asian third world countries in 14 Asian newspapers in 1977 and found that three-quarters of the news stories were supplied by the four news agencies (AP, UPI, Reuter, and AFP).

Studies on agenda-setting had been predominantly about media’s influence on individual audience, but, recently, ‘media agenda-setting’ is attracting more attentions of media scholars (McCombs, 2002). According to McCombs, media agenda-setting is a very complicated process in which political and economic power, PR strategies, and journalists’ tradition, practices, and professional values involve. However, it can be less complicated when the international news agenda-setting process is concerned since most of the newspapers are relying on news agencies for foreign news production and there are only three internationally influential news agencies (Reuters, AP and AFP) today. News
agencies are not only providing news items that are submitted on newspapers, but their news coverage also influences allocation of newspapers' journalistic resources for news production (Boyd-Barrett, 1980; Paterson, 2001): if news agencies are covering a certain event extensively, newspaper editors will try to dispatch journalists to produce their own reports, for example. Stevenson and Shaw (1984) found that there are strong correlations in international news coverage between newswire from Western news agencies (AP, UPI, AFP and Reuters) and local newspapers in Mexico, Argentina and Brazil; they argue that news agencies materials influence local editors and have agenda-setting effects on newspapers. According to Wu (2000), primary and secondary predictors of the amount of international news coverage are trade relationship between countries and existence of the international news agency bureaus. As McCombs (2004), in the context of domestic media agenda-setting, suggests that ‘subsidized’ news materials from government agencies is the one of the most important factors in media agenda-setting process, external provision of handy news items strongly influence news coverage.

Since the NWICO discussion in the 1960s, it has been argued that underrepresentation of the third-world countries in international news, as a result of the dominance of the Western news agencies, limits their possibility of economic growth and cultural independence. Recently, representation of developing countries is criticized for the reproduction of skewed image. On Western media, developing countries are described in stereotypical ways, so even if developing countries are covered by media, accurate report on complicated socio-economic problems of developing countries is missing (Hamelink, 2011; Miller, 2007; Philo, 2004; Thussu, 2004).

The decline in the international news is a long-term trend since the 1970s (J. M. Hamilton & Jenner, 2004). These days, news organizations often stop or reduce production of international news by themselves and start submitting news stories provided by news agencies instead, because many of them are facing strong commercial pressure (Freedman, 2009) and managements believe that news audience are not interested in international news (Arnett, 1998; Gorney, 2002). Hamilton (2006) found a lower level of interest in international affairs among 18-34 age-group females, whom are the most important target of advertisers, but the level of interests in foreign news among news audience is still under discussion.

Heavy reliance on news agencies is typical of online news services, which are rarely making profits and oftern lack resources to produce original news contents. Paterson (2007) reports that reliance on news agencies had risen from 68% in
2001 to 85% in 2006, although the small size of data used in the research demands careful interpretation of the result. News coverage on the most popular online news site, Yahoo! News and Google News, are largely determined by news agencies’ coverage: Yahoo! News only publishes international news from news agencies; Google News reflects news coverage on hundreds of news sites across the world, which hardly have capability to produce original international news stories and heavily rely on news agencies (Watanabe, 2013).

Research Questions

The international news gathering and distribution system has a hierarchy structure, through which global agenda is being set. Here, ‘global agenda’ refers to events or issues that are considered to be important by a large number of people across the world. It is very difficult for us to know the changes in global agenda but the process of the global agenda-setting can be recognized by analyzing international news flows. Usually, international news are produced and circulated by Western news agencies to national newspapers, and then delivered to news consumers, so news coverage by news agencies are expected to have agenda-setting effect on world population. However, to date, the level of the influence of news agencies on the international news coverage by news outlets across the world has not been statistically measured.

Considering the hierarchical structure of the international news flows, this research will address four types of news providers to grasp the global agenda-setting process: international news agencies (e.g. Reuters, AP and AFP), elite newspapers (e.g. The New York Times, The Washington Post, The Guardian, The Times, The Wall Street Journal, The Financial Times etc.), national newspapers, and news portal sites (e.g. Yahoo! News and Google News), so the first research questions is:

**RQ1**: How much do international news agencies have influence on the news coverage on newspapers?

It is news agencies which national newspapers are usually subscribing to but some elite newspapers are providing news to national newspapers. Furthermore, it is known that news editors are frequently checking other newspapers online edition in deciding news coverage (McCombs, 2004; Quandt, 2008), thus second research question is:

**RQ2**: If elite newspapers have influence on the news coverage on national
newspapers just as news agencies?

In recent years more and more people have come to read news online. We are exposed to international news not only on newspapers’ online editions but also on portal sites’ news services, which do not produce news themselves, so they are also becoming an important source of news these days. Therefore the third question is:

RQ3: Which type of news providers has the strongest influence on the news coverage on news portal sites?

Hypotheses

As Boyd-Barrett states “less than two dozen newspapers around the entire world could make a reasonable claim to independence in the gathering of a comprehensive international news” (Boyd-Barrett, 1980), most newspapers are depending on news agencies for foreign news gathering, there are handful of elite newspapers that are equipped with foreign correspondents and relying on news agencies to a lower degree. Therefore the first hypothesis is:

H1: Elite newspapers are less influenced by news agencies than national newspapers in determining news coverage.

Then, national newspapers are supplied with news stories by elite newspapers or their editors are consulting to elite newspapers websites, so national newspapers’ coverage is influenced by elite newspapers’ coverage but to a lesser extent than by that of news agencies:

H2: National newspapers’ coverage is influenced by elite newspapers but less than by news agencies.

Among news portal sites, there are two kinds of news services: ‘news redistributors’ that submit news items provided from non-online sources, and ‘news aggregators’ that resubmit news collected online. News redistributors are usually provided with news items by news agencies, so their coverage can be largely determined by news agencies. News aggregators collect news items from wider variety of sources including online editions of national newspapers, and their coverage is concentrated to country of major news events (Watanabe, 2013), so the coverage patterns are similar to the coverage of national newspapers news:
H3: News coverage by news redistributors is influenced by news agencies, while news the coverage by aggregators is influenced by national newspapers.

Methodology

In this research project, online editions of respective news providers are used as source of data to collect news items at small intervals for a long period of time to examine news flows. News items are collected from respective websites for one year from September 1st, 2011. Collected news items are coded according to their geographic coverage (country or region) using a content analysis program. The coding is aggregated and then statistically analyzed.

Case selection

As the international news agencies, Reuters and AP are chosen, but AFP is not included since it does not have own online outlet. The elite newspapers are The New York Times, The Washington Post and The Guardian, but The Times cannot be included because it restricts access to its website from non-subscribers, and The Wall Street Journal and The Financial Times are excluded as economic newspapers.

As national newspapers, eight English language newspapers are chosen from different parts of the world to observe a world-wide news dissemination process: The Times of India and The Hindu (India), The Straits Times (Singapore), The Times (South Africa), Daily Nation (Kenya), The Australian and The Sydney Morning Herald (Australia), and The New Zealand Herald (New Zealand). The reason only English newspapers are chosen was to exclude time-lags caused by translation works from English to local languages.

As news portal sites, news section of the most popular to portal sites, Yahoo! News and Google News are included. They respectively represent news redistributors and news aggregators.

Data collection

News items are collected from the online news outlets from above news providers. News on the online editions are not exactly the same as those on the printed edition but Stevenson (1994) and Wu (2007) report that there are only a small difference between them regarding international news coverage. Considering recent studies on online news, there may be more articles provided by
news agencies on online editions, but online news items are still chosen by news editors of newspaper publishes, so the tendency of the news coverage on online editions is assumed to be the same as on offline editions. The advantage using online news items is not only the low data collection cost but the irreplaceable quality that it allows us to sample news items very frequently. Today, news stories are disseminated across the world immediately after an event occurs and online news outlets are updated around the clock, so news sampling from print editions of newspapers, which are published at most twice a day, is not appropriate media to observe the speed of international news flows.

From the websites of the above mentioned newspapers and portal sites, only international news sections are chosen and their RSS feeds are registered to a news collection program. Between September 1st, 2011 and August 31st, 2012, the program had been monitoring the update of the RSS feeds every 10 minutes and collecting all new items. From the RSS entries, values in title and description fields are extracted and saved to a database with a timestamp. Most of the RSS feeds only provide summaries of news stories that are less than 50 words long, but they are sufficient for the purpose of this study. In addition to the news items from the newspapers and portal sites, many more items are collected just to construct a dictionary for a content analysis program explained below.

Content analysis

Content analysis of news items was conducted in terms of geographic focus of news stories to compare patterns of news coverage. In order to categorize a large number of news items, an automated content-analysis program1 was created from scratch by the author. Use of computer program for research in media studies is not new (e.g. Stone, Dunphy, Smith, & Ogilvie, 1966), but automated creation of a large dictionary is the distinctive feature of the program. Most of the analyses of large data sets in media studies have been conducted by keyword search (e.g. DFID, 2000; Eisensee & Strömberg, 2007; Golan, 2008; Miller, 2007; Van Belle, 2008), but Koler (2008) points out that imperfect knowledge about the media contents and inadequate keyword choice leads to unrepresentative sampling, because the contents retrieved from database are depending on the search words.

Thanks to a computer-generated large dictionary, the program can categorize approximately 90% of news items and inter-coder agreement with a human coder was above 80%2. The proportion of categorized items by the program is about 10%

1 See http://internationalnewsmap.info for demonstration.
2 The accuracy of the coding was measured prior to this research project using manual coding of randomly
higher than computer content analysis that only use a manually compiled dictionary (see Watanabe, 2013). Considering the inter-coder agreement on manual content analyses in terms geographic focus of news in previous research projects: 76.6% (Whitney, Fritzler, Jones, Mazzarella, & Rakow, 1989) and 80.1% (Jones, 2008), the categorization by the program is sufficiently accurate.

The content analysis program is created using PHP and MySQL database and it can categorize 10,000 news items in approximately 10 minute on an ordinary laptop PC purchased in 2010. The program is composed of three processes: (1) proper noun recognition; (2) dictionary creation; and (3) news categorization. The proper noun recognition process is implemented in the simplest but the most flexible way: if a word is more often found to be capitalized forms in the news items collected, it is considered to be a proper noun. The advantages of this algorithm are no need of prior knowledge about proper names and the high processing speed.

After the proper noun recognition, a large dictionary (called ‘dynamic dictionary’) is created from a small manually compiled seed dictionary (called ‘static dictionary’), which is composed of mostly names of places3. This process utilizes the unique structure of news items: in news items essential information such as origin of the news is shown in the headlines and more detailed information is contained in the body texts, so proper nouns for the dynamic dictionary can be extracted from body texts by categorizing only headlines using the static dictionary.

Categorization of headlines using the static dictionary (‘pre-coding’) is followed by collection of proper nouns from body texts4 for the dynamic dictionary. Collected proper nouns are given scores, which indicate the importance of the words to respective categories. Scores are calculated based on the commonality of a word within a category and across the all categories as described below. By the algorithm, unique words that appear only in news about specific countries (e.g. ‘Mubarak’) gain very high scores, but common words (e.g. ‘President’) are given very low scores.

This algorithm is inspired by the ‘word-score’ technique developed by Benoit

---

3 If we use names of famous people or organizations, the dictionary needs to be updated frequently, but names of places usually do not change. For example, words for Egypt are Egypt, Egyptian, Egyptians, Alexandria, Alexandrian, Alexandrines, Cairo, Cairene, and Cairenes. The complete dictionary is available online: http://www.pluri.info/site/resource/dictionary.txt

4 News items collected during last 7 days are used to create a dynamic dictionary.
and Laver (2003), so a score is gained by simply transforming tf-idf:

\[ S_{kc} = \log \left( \frac{\text{TF}(w_k|\hat{D}_c) \cdot \frac{1}{\text{DF}(w_k|\hat{D})} \cdot 100}{\text{DF}(w_k|\hat{D})} \right) \]

where \( S_{kc} \) is the score of a word, \( W_k \), for a country, \( c \), and \( \hat{D}_c \) is a set of items pre-coded as covering a country, \( c \). Here, in calculating a document frequency, a set of items pre-coded as the same country is treated as one ‘document’. The reason of the logarithm transformation is that some of the keywords have extremely high tf-idfs. Multiplication by 100 is simply to avoid gaining negative scores after the logarithm transformation.

Categorization is performed using the dynamic dictionary and words in the dictionary are searched for in the body texts of news items. Every time word in the dynamic dictionary is found in a body text, the score of the word is given to a corresponding country and the news item will be categorized as a country with the highest total score.

More precisely, where \( \text{TF}(W_{kc}|d_i) \) is a frequency of a word, \( W_{kc} \), associated with a country, \( c \), in a news item, \( d_i \), and \( S_{kc} \) is a score of the word, \( W_{kc} \), a total score of the country, \( S_c \), is:

\[ S_c = \sum_{k}^{n} \text{TF}(W_{kc}|d_i) \cdot S_{kc} \]

Categorization of items is performed by finding a country, \( c \), with a maximum score, \( S_c \). However, the scores of words are calculated by tf-idfs, so words relating to countries that are frequently covered by news prone to have lower scores. In order to offset this systematic error, scores need to be weighted by probability that the country appears among the news items, \( P(c|D) \). Nevertheless, we only have the result from pre-coding at this stage, so \( P(c|\hat{D}) \) is used instead:

\[ C = \arg \max_c \left( \sum_{k}^{n} \text{TF}(W_{kc}|d_i) \cdot S_{kc} \cdot P(c|\hat{D}) \right). \]

For example, if the following text is given:

---

\(^5\) term frequency–inverse document frequency
Mubarak Judge Summons Top Military Officials  The judge hearing the criminal trial of former President Hosni Mubarak summoned Field Marshal Mohamed Hussein Tantawi Gen Sami Anan and former intelligence chief Omar Suleiman (Kirkpatrick & Afify, 2011).

Egypt gets score of 216.36, Mauritania 6.25, Colombia 5.67, and Honduras 5.26 (Table1). The score of Egypt is obviously the highest, so this news item is categorized as such. What noteworthy is that none of the words ‘Mubarak’, ‘Hosni’, ‘Tantawi’, ‘Marshal’, ‘President’, and ‘Sami’, is included in the static dictionary for Egypt, but they are appropriately added to the dynamic dictionary by the program with very high scores. It is probably possible for researchers who compile dictionaries manually to include the name of the former president, Hosni Mubarak, but difficult to include the name of the Field Marshal, Mohamed Hussein Tantawi, which appeared frequently in news about Egypt only during the period.

Table 1: An example of scoring

<table>
<thead>
<tr>
<th>Country</th>
<th>Word</th>
<th>Transformed tf-idf</th>
<th>Weight</th>
<th>Count</th>
<th>Score</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egypt</td>
<td>Mubarak</td>
<td>301.89</td>
<td>0.33</td>
<td>2</td>
<td>200.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hosni</td>
<td>34.76</td>
<td>0.33</td>
<td>1</td>
<td>11.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tantawi</td>
<td>9.93</td>
<td>0.33</td>
<td>1</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marshal</td>
<td>1.99</td>
<td>0.33</td>
<td>1</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>President</td>
<td>1.72</td>
<td>0.33</td>
<td>1</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sami</td>
<td>0.99</td>
<td>0.33</td>
<td>1</td>
<td>0.33</td>
<td>216.36</td>
</tr>
<tr>
<td>Mauritania</td>
<td>Mohamed</td>
<td>1.49</td>
<td>4.20</td>
<td>1</td>
<td>6.25</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>Gen</td>
<td>1.40</td>
<td>2.74</td>
<td>1</td>
<td>3.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>President</td>
<td>0.66</td>
<td>2.74</td>
<td>1</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>Honduras</td>
<td>President</td>
<td>2.51</td>
<td>2.10</td>
<td>1</td>
<td>5.26</td>
<td></td>
</tr>
<tr>
<td>Somalia</td>
<td>Mohamed</td>
<td>2.41</td>
<td>1.89</td>
<td>1</td>
<td>4.55</td>
<td></td>
</tr>
<tr>
<td>Syria</td>
<td>President</td>
<td>0.89</td>
<td>3.50</td>
<td>1</td>
<td>3.12</td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>President</td>
<td>0.94</td>
<td>2.80</td>
<td>1</td>
<td>2.63</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>Gen</td>
<td>0.60</td>
<td>3.50</td>
<td>1</td>
<td>2.08</td>
<td></td>
</tr>
<tr>
<td>Cuba</td>
<td>Gen</td>
<td>2.23</td>
<td>0.72</td>
<td>1</td>
<td>1.61</td>
<td></td>
</tr>
<tr>
<td>Cameroon</td>
<td>President</td>
<td>1.25</td>
<td>1.20</td>
<td>1</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>Yemen</td>
<td>President</td>
<td>2.10</td>
<td>0.55</td>
<td>1</td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td>Rwanda</td>
<td>President</td>
<td>0.71</td>
<td>1.60</td>
<td>1</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>Malawi</td>
<td>President</td>
<td>1.37</td>
<td>0.70</td>
<td>1</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>Kiribati</td>
<td>President</td>
<td>0.50</td>
<td>1.75</td>
<td>1</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>President</td>
<td>1.33</td>
<td>0.66</td>
<td>1</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Venezuela</td>
<td>President</td>
<td>1.13</td>
<td>0.64</td>
<td>1</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>Gen</td>
<td>1.59</td>
<td>0.35</td>
<td>1</td>
<td>0.56</td>
<td></td>
</tr>
</tbody>
</table>

11
Measurement

Traditionally research on the influence of the news agencies have been conducted by studying bylines of news articles in international news sections on newspapers (e.g. Galtung & Ruge, 1965; Schramm, 1978) or, more recently, by finding concordance between news stories distributed by news agencies and those submitted on newspapers (e.g. Gaizauskas et al., 2001). The traditional method provides the most reliable measurement, but it cannot take into account increase in news production by newspapers’ journalists as a result of extensive coverage by news agencies; the relatively new method offers more flexible measurement, but it fails to identify news sources accurately when news stories are modified by newswire editors. Therefore, in this study, alternatively, only the patterns of geographic news coverage will be compared to capture the multiple dimensions that include news provision from news agencies, importance of news event perceived by newspaper editors, and newspaper publishers’ decision to dispatch own journalists.

There is no standard way to compare news coverage, so new methods need to be created. In this research project, patterns of news coverage are compared in two ways: by rank correlation and cross-correlation. The former method is used to compare entire patterns of coverage by news providers within a given period of time; numbers of news items covering respective countries are counted and compared by using Spearman’s rank correlation coefficient. A high rank correlation coefficient indicates similarity in news coverage patterns between given two news outlets. The cross-correlation analysis is used to identify time-lags of a certain country to be covered by news from two different news providers; news items covering the country are counted by every hour and moving averages of plus-minus 3 hours are calculated throughout the period (e.g. Chart 1), and then two waves of the moving average were compared to find a time lag between a pair of the waves that entail a maximum correlation coefficient (e.g. Chart 2). It is assumed that time-lags and similarity in coverage patterns indicate respectively direction and size of agenda-setting effects; relatively large time-lags between one news outlet to another is seen as an indication of a lower level of editorial autonomy, because, the news outlet is depending on an external news feed or its editorial judgment is influence by other news media. News editors are known to have herd mentality (Kaplan, Mohamedou, & Carver, 2002; McCombs, 2004; Quandt, 2008) and the time-lags in the editorial decision can be explained by the “threshold models of collective behavior” proposed by Granovetter (1978). He argues that each actor has a different level of threshold to initiate an action; some individuals initiate actions regardless of others while some actors wait until
majority of actors do so. If a news organization is capable of producing news and evaluating importance of a news event autonomously, it can publish news earlier than other organizations; but, if a news organization is depending on news feeds and influenced by news coverage by other organizations, it only can publish news later than others.

Cross-correlation analysis is a traditional method in research on agenda-setting. McCombs and Shaw (1972) had used cross-correlation analysis to determine causal relation in agenda-setting more than three decades ago. McCombs (1975) argued that when a variable X changes prior to a variable Y and those two variables strongly correlate, causal relation between X and Y can be inferred. However, he pointed out that, even in that case, a possibility that a variable Z that influence both X and Y to exist remains, so he urged an importance of theoretical grounds in claiming causal relations: when a causal relation between two variables is claimed, there should not be a rival hypothesis that suggests an existence of the third variable. The key differences in cross-correlation analysis used in this study and those used by McCombs are the frequency of the data collection and the size of time-lags: online news items are collected every 10 minutes for this research project thanks to the Internet and time-lag is set to be plus-minus 24 hours considering the speed of news flows on the Internet.

Some readers may question the accuracy of the time-lags observed on the websites, because it is not difficult to imagine that the news agencies submit news stories to organizational subscribers before they are published online to protect the value of their expensive services. However, even if this is the case, it only means that news agencies send news stories no later than they are published on website and the time difference only makes observed time-lags smaller than actual time-lags (this cannot be directly observed) and makes statistical analysis difficult.6

Results

During the data collection period, 299,822 items were collected from the news outlets. Among collected news items, those with a score of below 50 were excluded for confident coding and those covering a home country of the news provider were also eliminated as domestic news, so in total 227,532 items were used in the analysis (Table 2). The items covering Syria and China were found to be occupying

---

6 If observed time-lags are smaller than actually they are, it will be difficult to find statistically significant time-lags and to identify direction of causal effects between the news outlets.
a large proportion of news coverage and among the top 5 most covered countries by all the news providers. Egypt, Afghanistan and Iran were also one of the biggest origins of news events. The US and the UK did not appear often in the top 5 lists, but it is because most news items covering those countries had been eliminated as domestic news. News agencies seemed to be focusing on politically important countries that sometimes have conflicts, while newspapers are reporting economically important countries such as the US, China and India. The news coverage on Yahoo! News and Google News appeared to be between news agencies and newspapers. Gini coefficients were calculated to compare the degrees of concentration of news coverage and found that Google news has the most and news agencies has the least concentrated coverage patterns, as predicted by a previous study (Watanabe, 2013).

Table 2: News coverage distribution and numbers of items

<table>
<thead>
<tr>
<th></th>
<th>News agencies</th>
<th>Elite newspapers</th>
<th>National newspapers</th>
<th>Yahoo News</th>
<th>Google News</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>20,069</td>
<td>52,597</td>
<td>53,420</td>
<td>24,012</td>
<td>77,434</td>
</tr>
<tr>
<td>Top5</td>
<td>Syria (4.9%)</td>
<td>US (13.8%)</td>
<td>US (11.2%)</td>
<td>Syria (8.0%)</td>
<td>Syria (15.7%)</td>
</tr>
<tr>
<td></td>
<td>Iran (4.7%)</td>
<td>Syria (5.6%)</td>
<td>Syria (7.2%)</td>
<td>Iran (5.0%)</td>
<td>Iran (6.5%)</td>
</tr>
<tr>
<td></td>
<td>China (4.6%)</td>
<td>China (5.2%)</td>
<td>UK (5.8%)</td>
<td>Afghanistan (4.6%)</td>
<td>Egypt (5.7%)</td>
</tr>
<tr>
<td></td>
<td>Egypt (4.0%)</td>
<td>India (4.7%)</td>
<td>China (5.5%)</td>
<td>China (4.5%)</td>
<td>Afghanistan (5.4%)</td>
</tr>
<tr>
<td></td>
<td>Russia (3.6%)</td>
<td>Afghanistan (3.9%)</td>
<td>India (4.3%)</td>
<td>Egypt (4.4%)</td>
<td>China (4.4%)</td>
</tr>
<tr>
<td>Gini</td>
<td>0.78</td>
<td>0.83</td>
<td>0.85</td>
<td>0.81</td>
<td>0.87</td>
</tr>
</tbody>
</table>

In order to compare the patterns of news coverage in a more scientific manner, numbers of news items were summed by country and were compared using mean values of Spearman’s rank correlation coefficients calculated for each day (Table 3). The reason mean values of daily rank correlation coefficients were used to compare news coverage patterns instead of rank correlation coefficients of whole period is that most covered countries are almost the same in all types of news providers when the data is aggregated. The rank correlation coefficients show that the coverage on national newspapers is more similar to that of elite newspapers than news agencies; Yahoo! News’ coverage is more like news agencies’ coverage than newspapers’ coverage. Yahoo! News’ relatively large standard deviation of the correlation coefficient with news agencies is due to the changes in Yahoo! News’ coverage pattern during the data collection period: the mean value of the correlation coefficients had been 0.45 until February 5th, 2012, but it has become 0.88 after the day. The coefficients of Google News with all other types of news providers are very close, but Tukey’s tests at 95% confidence level indicated the correlation with national newspapers is significantly higher than others.
The similarity of news coverage patterns is very informative but it does not indicate any direction of news flows, so cross-correlation analysis was performed next. For the analysis, a moving time window of plus-minus 3 hours for a given time was defined and numbers of items covering a certain country within the time window were computed (e.g. Chart 1); the expansion of the time window was to avoid accidental high correlation coefficients. Then, using a pair of such data, a time-lag that maximizes correlation coefficients was searched for between plus-minus 24 hours using a cross-correlation function in R program (e.g. Chart 2). This analysis was repeated for all 240 countries and for all combinations of the news providers to collect time-lags with maximum correlation coefficients.
When there is a systematic time-lag between two news providers, the mean values of time-lags are supposed to be statistically significantly different from zero\(^7\), so statistical tests were performed by using two-tailed t-test (Table 4). The coverage by news agencies precedes elite newspapers 1.18 hours, national newspapers 2.77 hours and Google News 4.95 hours. The difference between news agencies and Yahoo! News was not statistically significant: the distribution of their time-lags were far from normal distribution\(^8\). Elite newspapers do not have succeeding news providers but national newspapers are succeeded by Google News with 1.54 hour time-lag. Yahoo! News precedes elite newspapers 2.31 hours, national newspapers 2.08 hours and Google News 3.50 hours, but the difference in time-lags between elite and national newspapers was not statistically significant.

Table 4: Mean values of time-lags with maximum cross-correlation coefficients

<table>
<thead>
<tr>
<th>News agencies</th>
<th>Elite newspapers</th>
<th>National newspapers</th>
<th>Yahoo! News</th>
<th>Google News</th>
</tr>
</thead>
<tbody>
<tr>
<td>News agencies</td>
<td>-1.18 (0.44)*</td>
<td>-2.77 (0.51)***</td>
<td>-0.46 (0.19)</td>
<td>-4.95 (0.71)***</td>
</tr>
<tr>
<td>Elite newspapers</td>
<td>1.18 (0.44)*</td>
<td>-0.02 (0.32)</td>
<td>2.31 (0.56)***</td>
<td>-1.09 (0.63)</td>
</tr>
<tr>
<td>National newspapers</td>
<td>2.77 (0.51)***</td>
<td>0.02 (0.32)</td>
<td>2.08 (0.45)***</td>
<td>-1.54 (0.49)*</td>
</tr>
<tr>
<td>Yahoo! News</td>
<td>0.46 (0.19)</td>
<td>-2.31 (0.56)***</td>
<td>-2.08 (0.45)***</td>
<td>-3.50 (0.71)***</td>
</tr>
<tr>
<td>Google News</td>
<td>4.95 (0.71)***</td>
<td>1.09 (0.63)</td>
<td>1.54 (0.49)*</td>
<td>3.50 (0.71)***</td>
</tr>
</tbody>
</table>

N=240; Standard errors reported in brackets; Signif. codes: 0 **** 0.001 *** 0.01 ** 0.05 *. 0.1 *. 1.

Elite and national newspapers demonstrated very similar patterns in news

---

\(^7\) If there is no systematic time-lag between two news providers, the time-lags with maximum correlation coefficients distribute randomly, so their mean value becomes very close to zero.

\(^8\) Time-lags with maximum correlation coefficients were zero in 220 cases.
coverage: the Gini coefficients, which indicate concentration of news coverage, were very close; the mean values of rank correlation coefficients with news agency were the same level (0.57) and it was 0.70 between them; elite newspapers had a smaller mean value of maximum cross-correlation coefficients (0.24) than national newspapers (0.27) with news agencies, but it was not a statistically significant difference (p=0.09) (Table 5). The only major difference elite and national newspapers was time-lags with news agencies (-1.18hr and -2.77hr); a result of a t-test indicated that the difference between the time-lags was statistically significant (p=0.02). Given the similarity in patterns of elite and national newspapers, we have to doubt that there is a substantial difference in level of influence from news agencies between them, so the first hypothesis (H1) was not supported. The difference in time-lags with news agencies between them can be interpreted that elite newspapers influences national newspapers but both are equally influenced by news agencies.

Table 5: Mean values of maximum cross-correlation coefficients

<table>
<thead>
<tr>
<th>News agencies</th>
<th>Yahoo! News</th>
<th>Google News</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elite newspapers</td>
<td>0.24 (0.15) 0.27 (0.16)</td>
<td>0.60 (0.19) 0.26 (0.14)</td>
</tr>
<tr>
<td>National newspapers</td>
<td>0.40 (0.18) 0.21 (0.14)</td>
<td>0.29 (0.14)</td>
</tr>
<tr>
<td>Yahoo! News</td>
<td>0.29 (0.17)</td>
<td>0.37 (0.18)</td>
</tr>
<tr>
<td>Google News</td>
<td></td>
<td>0.28 (0.18)</td>
</tr>
</tbody>
</table>

N=240; Standard deviations reported in brackets.

The influence of the elite newspapers on national newspapers is difficult to detect, because both elite and national newspapers are strongly influenced by news agencies. The time-lags between elite on national newspapers appeared to be very small and insignificant, so it is not certain whether the changes in news coverage on national newspapers are caused by news agencies or elite newspapers based on above analyses, but when the data was split by the median value of amount of news coverage by national newspapers (114), opposite time-lags were found: the mean values for countries above and below median were 0.47 and -1.01 and they are statistically significantly different according to ANOVA (p=0.02). These opposite time-lags mean that national newspapers are slightly preceding elite newspapers in covering major countries of news events, while they are still succeeding the elite newspapers in covering minor counties. Therefore, the second hypothesis (H2) was partially supported: elite newspapers have

---

9 Opposite time-lags were found only between elite and national newspaper.
influence on national newspapers in covering major countries but not in covering minor countries.

Finally, all results indicate that the news coverage by news redistributor (Yahoo! News) is more similar to news agencies, while that of news aggregator (Google News) is more like that of national newspapers: Yahoo! News and news agencies, respectively, have the first and the second smallest Gini coefficients (0.81 and 0.78), but Google News and national newspapers have the highest ones (0.87 and 0.85); mean values of rank correlation coefficients between news agencies and Yahoo! News (0.72) and between national newspapers and Google News (0.60) are among the highest, so the third hypothesis (H3) was strongly supported.

**Modeling**

The result of the cross-correlation analysis successfully estimated time-lags with which news is released by different types of news providers. Based on the estimated time-lags, we can understand the international flows of news more precisely and construct a model of international news flows (Chart 3). News agencies are at the top of the hierarchy of global news flows and provide all types of news outlets with news stories. Elite newspapers’ coverage succeeds 1.18 hours after that of news agencies. National newspapers’ coverage changes 2.77 hours after news agencies, but its coverage of minor countries is influenced by elite newspapers.

Google News’ coverage has statistically significant time-lags with news agencies (4.95), national newspapers (1.54) and Yahoo! News (3.50), but it seems that it is only national newspapers that directly influence the news coverage on Google News for three reasons: (1) it is collecting news from websites of national newspapers across the world (Watanabe, 2013); (2) the mean value of rank correlation coefficients with national newspapers is significantly higher than others; (3) the time-lag with news agencies is close to a total of time-lags (4.31) between news agencies and national newspapers (2.77) and between national newspapers (1.54) and Google News. Therefore, the news flows only from national newspapers to Google News are included in the mode.

Yahoo! News’ coverage precedes elite and national newspapers and Google News, but it was not considered to have influence on those news outlets, because it is just a redistributor of newswire. Its time-lags with news agencies with maximum correlation coefficients were zero in 220 cases, so it seems that it is connected to news agencies’ news distribution network directly and submits news
items immediately after the news agencies.

Chart 3: A model of international news flows

Discussion

The rank-correlation analysis produced results that endorse our previous knowledge: national newspapers are mostly covering news from powerful countries (Himelboim, Chang, & McCreery, 2010; Stevenson & Cole, 1984; Stevenson, 1994; Wu, 2007); Yahoo! News is heavily relaying on news agencies for news production (Palser, 2002; Watanabe, 2013); Google News’ coverage reflects coverage on small news outlets across the world (Watanabe, 2013). To date, there has not been a standard method to analyze international news flows, but the results demonstrate that rank correlation and cross-correlation analyses of news coverage are very useful methods in international news studies.

However, some results of the analysis contradict our knowledge. Our belief that elite newspapers have higher degree of autonomy is found to be questionable. They can release news in 1.18 hours after the news agencies’ release while national newspapers take 2.77 hours, but their patterns of news coverage are very similar, so they seem to be influenced by news agencies to the same degree. In other words, elite newspapers are more autonomous than national newspapers in a sense that they are less influenced by other newspapers, but not independent from news agencies.

The influence from elite newspapers to national newspapers were only limited to coverage on minor countries. The best possible account of this is that news editors of national newspapers are routinely choosing news about major countries
but careful about publishing news about minor countries; news stories about major countries are automatically considered to be important to the audience but news about major countries are not, so they need to wait until other newspapers, especially elite newspapers, to publish the news.

As stated in the methodology part, large time-lags were initially considered as a sign of a lack of editorial autonomy, but this does not always apply, as we see in Yahoo! News. Despite the fact that Yahoo! News is heavily relying on news agencies (Palser, 2002; Watanabe, 2013), the time-lag was less than one hour. It is probably because Yahoo! News’ backend system is connected to news agencies’ news distribution network and their feeds are instantly submitted to Yahoo! News’ website without any human involvement in news selection. According to Granovetter’s model of collective action, actors are supposed to have some degrees of autonomy and can decide whether join a collective action or not, so the action spreads with time-lags. However, if there is no autonomy in actors at all, the collective action spreads immediately without a time-lag. This seems to be the case of Yahoo! News, so the size of time-lags indicated levels of agenda-setting effects only when news outlets have a certain degree of autonomy in editoriala decision.

As we can see in the brackets next to types of news providers in the chart (Chart 3), the Gini coefficients increase as news flows down to the lower tiers: the Gini coefficient is 0.78 initially but becomes 0.87 in the end. Conventionally, concentrated news coverage to major countries has been understood as a result of the dominance of Western news agencies since a NWICO debate, but the result shows a more complicated picture: the concentration is not only due to the news agencies’ Western focus but also the multiplications of skewed news coverage distribution through the cascade-like global agenda-setting process. This can be explained by the herd mentality of news editors and Granovetter’s model of collective behavior: news coverage on a news outlet is determined by amount of news items published on other news outlet previously and news items are published only when the amount of news items published on other news outlets exceeds a certain threshold, but amount of news about minor countries does not exceeds the threshold often, so the proportion of news about major countries continue to increase.

Overall, the lagged increase in numbers of news items and the progressive increase in concentration of news coverage in different types of news providers indicate an existence of a network of news publishers. News publishers are diverse in quality and by no mean connected by official institutional ties, but their editorial choice is visible to each other on the website and one’s editorial choice is depending on others’ editorial choice. In this sense, the relationship among the
news publishers is very similar to Granovetter’s individuals who share the same physical space and gradually join a collective action. The idea of the network of news publishers allow us to apply theories of collective action to media agenda-setting process and provide us with better understanding of the global media agenda-setting process.

**Contribution**

As methodological a contribution, a highly accurate news categorization program was created and its algorithm was explained in detail to allow other researches to develop their programs. The news categorization program allows individual or small research groups to carry out a large scale analysis of news.

The most interesting finding in this research project is the possibility to estimate time-lags with which news flows internationally using the cross-correlation analysis. Based on the estimated time-lags, news flows between different types of news providers could be identified and a model of global news flow could be created. To the author’s knowledge, international news flows have never been measured with time-lags.

The outcomes of this research bridge widely recognized classic theories of collective behavior (Granovetter, 1978) and agenda-setting (McCombs & Shaw, 1972; McCombs, 1975) and recently developing fields, such as international communication and online media, and contribute to expand the horizon of media studies.

**Reference**


Americans’ Policy Preferences (1st ed.). University Of Chicago Press.


