Letters to the editor

Text mining analysis of newspaper editorials concerning the COVID-19 pandemic from a healthcare perspective

Wakae Maeda¹, Yoshihisa Hirakawa², Tsukasa Muraya³, and Hisayuki Miura⁴

¹Faculty of Nursing, Chukyo Gakuin University, Japan
²Department of Public Health and Health Systems, Nagoya University Graduate School of Medicine, Japan
³Faculty of Design, Kyushu University, Japan
⁴Department of Home Care and Regional Liaison Promotion, National Center for Geriatrics and Gerontology, Japan

Abstract

Objective: This pilot study aimed to examine the content of Japanese newspaper editorials concerning the coronavirus disease 2019 (COVID-19) pandemic and its change over time using text mining analysis.

Materials and Methods: The authors analyzed qualitative data from the editorials of five national and 12 regional newspapers on April 7 and 8, 2020 (first state of emergency) and January 8, 2021 (second state of emergency). All analyses were conducted using KH Coder version 3.

Results: The co-occurrence network showed a low level of content diversity and a high degree of politicization in the COVID-19 news coverage. The top five high frequency words from the newspapers were “infection”, “declaration”, “healthcare”, “government”, and “emergency” at the first state of emergency, and were “declaration”, “measures”, “government”, and “restaurant” at the second one.

Conclusion: The results suggest a lack of detailed information and recommendations concerning the public health challenges of the COVID-19 pandemic in Japanese newspaper editorials, even one year after the first wave of the pandemic. This study provides a data-driven foundation for the effectiveness of newspapers in COVID-19 public health communications. The extent to which the quantity and quality of information from newly emerging communication channels, such as social media, influences public understanding of public health measures remains to be established.

Key words: coronavirus disease 2019 (COVID-19), healthcare, newspaper, state of emergency, text mining

Introduction

The impact of the coronavirus disease 2019 (COVID-19) pandemic on the Japanese economy and health systems is evolving in real time. Japan declared a state of emergency in Tokyo and other prefectures where the number of confirmed cases and the risk of infection were high, and went into a first “mild” lockdown early in the pandemic (April 7, 2020). Japan-style lockdowns carry fewer restrictions on daily life and business activity than in European countries, including France, Italy, and the UK¹,². In a state of emergency, the Japanese government attempted to reduce foot traffic by asking restaurants to shorten their hours and refrain from serving alcohol, and urging companies to allow staff to work from home more frequently¹,². During a state of emergency, a governor can advise local people to avoid unnecessary outings; however, residents have the right to refuse these requests, and there are no penalties for disobedience¹,². In the early stages of the third COVID-19 wave, the Japanese government declared a second state of emergency in Tokyo and surrounding areas (January 8, 2021), as COVID-19 infections and the number of people in serious condition reached record levels.

Thus, Japan’s version of a state emergency is request-
based and does not necessarily result in successful and sustained behavioral changes among residents. During the “mild” lockdown, the residents were unlikely to remain confined to their homes, and proactive decision-making was left up to individuals. Under the laws and regulations of Japan, there are no penalties for going out for non-essential reasons, and any change in the population’s behavior must be voluntary. Behavior change interventions to work effectively and deliver population-level behavioral outcomes must be underpinned by behavioral theory. The health belief model, originally developed in the 1950s, has served as a social psychological health behavior change model to explain and predict health-related behaviors. The model is based on the theory that a person’s willingness to change their health behaviors is primarily because of their perceptions toward the recommended behavior or action.

An editorial presents the newspaper’s opinion on an issue. Editorial writers argue and attempt to persuade readers to follow their beliefs. Editorials are written to influence public opinion, promote critical thinking, and sometimes encourage people to act on an issue. Although newspaper reading is declining among younger generations, who prefer to click on a news website on demand to read an article of their choice, newspaper editorials can still affect the majority vote of readers, reduce people’s anxiety toward an issue, and reassure people and society. Therefore, editorials should properly consider the general public’s level of comprehension of health-related issues, which in turn can have a significant impact on public health and related policies.

In Italy, Corvo and De Caro analyzed news coverage (printed press and online articles) during the COVID-19 pandemic/epidemic using data from 1,512 articles in two months. However, as in other countries, very few studies have investigated the detailed text of newspapers during the COVID-19 pandemic in Japan. More attention should be paid to analyzing the content of newspaper editorials, which could have an impact on public behavior and reactions. Moreover, partly because evidence of anti-COVID-19 measures, such as the use of face masks and disinfection methods by the public to impede COVID-19 transmission, is accumulating rapidly, the authors hypothesized that Japanese newspapers would attempt to update the content of their editorials in a timely manner to reflect the latest COVID-19 information. This pilot study aimed to examine the content of Japanese newspaper editorials and their changes over time during the COVID-19 pandemic.

## Results

The total number of sentences extracted from newspapers in the first state of emergency was 1,195, and the total number of paragraphs was 556.

The co-occurrence network generated from this analysis is shown in Figure 1. In this network, a node represents a high-frequency word in all articles. An edge represents the co-occurrence relationship between two high-frequency words appearing simultaneously in the subject of the same article. Nodes with a large degree are often considered high-connectivity or hub. The figure seems to be related to the government declaring a state of emergency in response to the novel coronavirus disease and closely monitoring the infection situation as well as the burden on hospitals. This result also seems to imply that the government asked the people to refrain from non-essential outings during the state

## Materials and Methods

This study involved text mining of editorials from national or regional daily print newspapers, and no human subjects were involved in the study; thus, ethics approval was not required. Text mining is the process of mining valuable insights from a large amount of unstructured text data. In recent years, studies using text mining have been commonly conducted in clinical fields, including public health.

The authors acquired the editorials of relevant newspapers on April 7 and 8, 2020 (first state of emergency) and January 8, 2021 (second state of emergency) from the websites of the largest news archive search engines in Japan, Google, and Yahoo: five national newspapers (Asahi, Mainichi, Nikkei, Sankei, and Yomiuri) and 12 regional newspapers (Hokkaido, Kahoku Shimpo, Niigata Nippo, Shinano Mainichi, Chunichi, Kyoto, Kobe, Chugoku, Sanin, Ehime, Nishinihan, and Okinawa) were selected considering the highest circulation number in each region.

The authors first imported the qualitative data as text data into a computer, then decomposed the dataset into words by morphological analysis by decomposing language into smaller segments, with morphemes being the smallest meaningful units of language having grammatical meaning, and classified the units into parts of speech. Second, after analyzing the trends and features of the classified morphemes, frequently occurring words and keywords were extracted and their frequency and simultaneous occurrence relationships were analyzed. The co-occurrence network diagram expresses the frequency of words and co-occurrence of two or more words. The diagram indicates stronger co-occurrence relationships by thicker solid lines and depicts words with a higher number of occurrences in larger circles.

The analysis was initially performed in Japanese, and the results were translated into English when preparing this research paper. The English language translation was verified by back-translation. Co-occurrence analysis and word count analysis were conducted using KH Coder version 3 (Kyoto, Japan) because it is a free, readily available, and widely used Japanese software originally developed in the fields of sociology and social research.
of emergency.

The words extracted by text mining for the first and second states of emergency and their frequencies of use are listed in Table 1. Five words, “infection”, “declaration”, “healthcare”, “government”, and “emergency”, ranked highest in terms of frequency of appearance. This seems to be related to the fact that these words were used to state that the government had declared a state of emergency in response to the novel coronavirus disease and to suggest that the government continued to closely assess the infection situation as well as the burden on the medical treatment structure. Among the other words that ranked in the top 50 were healthcare-related words, including “new”, “corona”, and “virus”. This result implies a newly emerged coronavirus disease (COVID-19).

The total number of sentences extracted from newspapers in the second state of emergency was 561, and the total number of paragraphs was 255. Five words, “infection”, “declaration”, “measures”, “government”, and “restaurant”, ranked the highest in terms of frequency of appearance. This result seems to have occurred for the same reason as for the first state of emergency and because restaurants and food-service businesses were among the most severely affected economic activities by the COVID-19 pandemic. Among other words that ranked in the top 50 were healthcare-related words, including “healthcare”, “corona”, “new”, and “virus”. The result was the same as that of the first state of emergency, implying the necessity of medical care for the newly emerged coronavirus.

Discussion

This study suggests a lack of detailed information and recommendations concerning public health challenges, such as infection prevention and control of COVID-19 among Japanese newspaper editorials, even one year after the first coronavirus pandemic. The results of the co-occurrence network showed a low level of content diversity and a high degree of politicization in COVID-19 news coverage. The results were similar to those of Corvo and De Caro’s study: Italian newspapers showed overall similar content, with more focus on the socio-political-economic impact of COVID-19 than on public health. This study provides a data-driven foundation for the use of traditional media such as newspapers in COVID-19 public health communication strategies.
There are several possible explanations for these unexpected results. First, because COVID-19 is a newly emerging disease, the authors of newspaper editorials most likely lacked satisfactory evidence-based healthcare literature to refer to. Second, in addition to the lack of relevant published literature on COVID-19, newspapers might have found it difficult to cover the topic in person because of social distancing. Third, because the virus poses a genuine threat to people’s daily lives and existence, the regional and local impact of the COVID-19 crisis was highly heterogeneous, with trade-offs between optimal health and economic risks. Therefore, the significance of the COVID-19 impact could have discouraged writers from exploring their opinions further. Finally, general news articles may have covered the research topic at length and represented a good source of information on people’s attitudes concerning the COVID-19 pandemic from a healthcare perspective. In addition, social media posts may have been a good source of information on people’s attitudes toward the pandemic. Further studies are needed to comprehensively collect the relevant qualitative data.

### Conclusion

This study examined the content of Japanese newspaper editorials and their changes over time using text-mining analysis. The results suggest a lack of detailed information and recommendations concerning the public health challenges of the COVID-19 pandemic among Japanese newspaper editorials, even one year after the first wave of the pandemic. This study provides a data-driven foundation for the effectiveness of newspapers in COVID-19 public health communications. The extent to which the quantity and quality of information from newly emerging communication channels, such as social media, influences public understanding of public health measures remains to be established.

### Acknowledgments

The authors disclose receipt of the following financial support for the research, authorship, and publication of this article. This work was supported by a Health Labor Sciences Research Grant [grant number 21GB1001].

### References

1. Yamamoto T, Uchiumi C, Suzuki N, et al. The psychological impact of ‘mild lockdown’ in Japan during the COVID-19 pandemic: a nationwide survey under a declared state of emergency. Int J Environ Res Public Health 2020; 17: 9382. [Medline] [CrossRef]
2. Watanabe T, Yabu T. Japan’s voluntary lockdown. PLoS One 2021; 16: e0252468. [Medline] [CrossRef]
3. Jones CL, Jensen JD, Scherr CL, et al. The Health Belief Model as an explanatory framework in communication research: exploring parallel, serial, and moderated mediation. Health Commun 2015; 30: 566–576. [Medline] [CrossRef]
4. Corvo E, De Caro W. COVID-19 and newspapers: a content & text mining analysis. Eur J Public Health 2020; 30: ckaa165.064. [CrossRef]
5. Higuchi K. A two-step approach to quantitative content analysis: KH coder tutorial using Anne of Green Gables (Part I). Ritsumeikan Soc Sci Rev 2016; 52: 77–91.

---

**Table 1** Word occurrence frequency in national and regional newspapers on April 7 and 8, 2020 (first state of emergency) and January 8, 2021 (second state of emergency)

| Rank | 2020 Frequency | 2021 Frequency |
|------|----------------|----------------|
| 1    | Infection      | 154            | Infection      | 166            |
| 2    | Declaration    | 101            | Declaration    | 76             |
| 3    | Healthcare     | 73             | Measures       | 71             |
| 4    | Government     | 56             | Government     | 68             |
| 5    | Emergency      | 48             | Restaurant     | 56             |
| 6    | Restriction    | 48             | Spread         | 49             |
| 7    | Situation      | 47             | Prime minister | 46             |
| 8    | People         | 47             | Healthcare     | 43             |
| 9    | Measures       | 47             | Request        | 42             |
| 10   | Subject        | 45             | Situation      | 40             |
| 11   | Spread         | 44             | Coronavirus    | 38             |
| 12   | Necessary      | 44             | Announcement   | 38             |
| 13   | Citizens       | 43             | Ask            | 37             |
| 14   | Request        | 40             | Citizens       | 34             |
| 15   | Outing         | 39             | Tokyo          | 34             |
| 16   | Prime minister | 39             | Emergency      | 33             |
| 17   | Economy        | 36             | Cooperation    | 32             |
| 18   | New            | 36             | Last year      | 30             |
| 19   | Coronavirus    | 35             | Business       | 29             |
| 20   | Action         | 34             | Suga           | 27             |
| 21   | Community      | 33             | Necessary      | 27             |
| 22   | Company        | 32             | Condition      | 26             |
| 23   | Ask            | 32             | New            | 26             |
| 24   | Self-controle  | 31             | Subject        | 26             |
| 25   | Tokyo          | 31             | Self-controle  | 25             |

| 45   | Avoid          | 20             | People         | 18             |
| 46   | Abe            | 19             | Based on       | 17             |
| 47   | Based on       | 19             | Period         | 17             |
| 48   | Institution    | 19             | Ramp-up        | 17             |
| 49   | Lifestyle      | 19             | Show           | 17             |
| 50   | Explanation    | 19             | Prevention     | 17             |