Citation:
Mykkänen, M and Patron, A and Badham, M and Valentini, C (2024) Leveraging NLP for crisis communication management: A case study of news media analysis of the COVID19 pandemic in two Nordic countries. Journal of Contingencies and Crisis Management, 32 (3). pp. 1-19. ISSN 0966-0879 DOI: https://doi.org/10.1111/1468-5973.12611

Link to Leeds Beckett Repository record:
https://eprints.leedsbeckett.ac.uk/id/eprint/11230/

Document Version:
Article (Published Version)

Creative Commons: Attribution-Noncommercial 4.0

© 2024 The Author(s)

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please contact us and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on openaccess@leedsbeckett.ac.uk and we will investigate on a case-by-case basis.

Check for updates

Received: 3 November 2023 | Revised: 4 July 2024 | Accepted: 2 August 2024

DOI: 10.1111/1468-5973.12611

ORIGINAL ARTICLE

WILEY

Leveraging NLP for crisis communication management: A case study of news media analysis of the COVID-19 pandemic in two Nordic countries

Markus Mykkänen[1] | Anri Patron[2] | Mark Badham[3] | Chiara Valentini[4]

[1]Department of Language and Communication Studies, University of Jyväskylä, Jyväskylä, Finland

[2]Department of Computer Science, University of Helsinki, Helsinki, Finland

[3]Leeds Business School, Leeds Beckett University, Leeds, UK

[4]Jyväskylä University School of Business and Economics, Jyväskylä, Finland

Correspondence
Markus Mykkänen, Department of Language and Communication Studies, University of Jyväskylä, Jyväskylä, Finland.
Email: markus.mykkanen@jyu.fi

Funding information
Helsingin Sanomat Foundation

Abstract

The news media plays a vital role in influencing public perceptions about topics, issues and crises. They also act as important intermediaries between organizations and public, enabling organizations to shape how people think about crisis topics and actors. Monitoring news coverage and assessing the news media's agenda-setting role in a crisis can help organizations respond more effectively to emerging situations. When crises are prolonged and affect many countries, media analysis can become a tedious task for crisis managers. This study demonstrates how natural language processing (NLP) methods can be utilized in news media analysis of crisis situations, such as an extended cross-national pandemic. Specifically, it demonstrates the possibilities of using NLP to identify and compare the salience of diverse crisis topics and how the media treat these topics and crisis actors (first- and second-level agenda-setting) across countries, news outlets, and time. The COVID-19 pandemic serves as an illustrative case study to showcase the application of NLP techniques to provide insights into public perceptions of a major health crisis shaped by the news media in two Nordic countries (Finland and Sweden). Findings show the suitability of NLP methods to detect nuanced differences in news media coverage and offer relevant knowledge of how public perceptions and responsibility attribution fluctuate across time and countries.

KEYWORDS
agenda-setting, COVID-19, crisis, natural language processing, news media

1 | INTRODUCTION

The news media constitutes a dominant force of public influence in society (Carroll & McCombs, 2003) and plays a particularly significant influencing role during major crises (Seeger et al., 2003), assisting public authorities to disseminate crisis information to the public, helping the public make sense of dynamic and complex crisis messages and offer their own interpretations of crisis messages (e.g., in editorials). The role of news media is crucial in crisis management due to its significant influence on public responses to a crisis and the extent by which instructive messages are followed (Coombs, 2021). Crises are perceptual (Coombs, 2012) and shaped by human interpretation (Gigliotti, 2020). To manage crises, a thorough understanding of how news media highlight, interpret and represent crisis

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2024 The Author(s). *Journal of Contingencies and Crisis Management* published by John Wiley & Sons Ltd.

situations, their key players, and events is vital for assigning crisis responsibility, assessing reputational risks and devising effective crisis management strategies.

Agenda-setting theory (McCombs & Shaw, 1972) provides valuable insights into the role of news media in crisis management, highlighting how media narratives not only influence public perceptions of the salience of crisis topics, but also the attribution of responsibility. The first level of agenda-setting (Carroll & McCombs, 2003; McCombs, 2004, 2005) illustrates the media's ability to shape public perceptions about what crisis topics matter most, while a second level of agenda-setting is revealed when journalists assign certain attributes to crisis stories and topics (Ghanem, 1997). A third level of agenda-setting examines how interrelationships among crisis topics covered by the media can be transferred from the media agenda to the public agenda (Guo & McCombs, 2011), thus highlighting the salience of the network agenda. Further, intermedia agenda-setting studies (Golan, 2006) have shown the agenda-setting effect between types of media (i.e., how some media outlets' coverage of topics influence other outlets' coverage of those topics).

In this study, we take the case of the COVID-19 pandemic to illustrate how natural language processing (NLP) methods, computer-assisted analytical techniques aimed at automatically analysing and comprehending human language (Manning & Schütze, 1999), can be used by crisis managers to understand the agenda-setting role of news media in two Nordic countries, Finland and Sweden. NLP methods are very useful tools to track, map and compare a large corpus of news media articles. They can provide important assistance in quickly generating crisis knowledge from analysing and sorting large sets of data, such as news reports, which would not be possible to achieve with other methods. The COVID-19 pandemic offers a unique opportunity to explore the potential of NLP applicability in a large data set and gain insights into the changes of agenda-setting processes shaping public perceptions of the COVID-19 pandemic across time, geographical space and media type. Specifically, we intend to explore whether NLP methods can be used to extract information on the first level (*what* is discussed) and second level (*how* it is discussed) of the news media's agenda-setting role. While the use of NLP methods in detecting agenda-setting patterns has been tested earlier, most of these studies focused only on the first level, and on single countries, often with English-speaking textual corpora (see, e.g., Boumans & Trilling, 2018; Chen et al., 2023; Hovy & Prabhumoye, 2021; Kim et al., 2017; Pang & Lee, 2008). Although it is generally acknowledged that agenda-setting is a complex and dynamic process and that agendas may fluctuate over time, very few studies have examined the time lag between agendas (Kim et al., 2017). This is particularly important when studying changes in and across agendas throughout a long-term situation or event.

To address these gaps, this study explores agenda-setting across more than 2 years of the COVID-19 pandemic in two Nordic countries (Finland and Sweden), whose language groups, Finnish and Swedish, are still largely understudied in the context of NLP applicability. This choice of focusing on two Nordic countries was purposeful, as the two countries offer both similarities and differences in authority-led pandemic management that could be valuable to consider when planning crisis response activities (Valentini & Badham, 2023). Furthermore, while these countries' media are consistently rated high in terms of media transparency and media trust (Andreasson, 2017), there are clear differences, and several different journalistic styles between Nordic journalists (Hovden & Väliverronen, 2021). Given that media systems vary across countries (Hallin & Mancini, 2004) and thus that the agenda-setting function may vary across countries, we argue that this study's comparative investigation of the media's agenda-setting role across these two countries contributes to a better understanding of multicountry differences in the way media shape the public agenda (McCombs, 2005) during a crisis.

2 | LITERATURE REVIEW

2.1 | News media role shaping publics' crisis perceptions

While the news media is one of the most dominant forces of public influence in most societies (Happer & Philo, 2013), the news media also enables organizations to shape their coverage of topics, issues and crises (Ali & Sherman-Morris, 2023; An & Gower, 2009; Öhman et al., 2016) and this shapes public perceptions about them. The extant crisis communication literature has constantly highlighted several central roles of news media in the crisis communication management process, from spreading public awareness about crisis-related issues to instructing and informing crisis guidelines, to playing a sense-making role amidst confusing and contradictory information (e.g., Boin et al., 2009; Bowen & Zheng, 2015; Kim, 2016; Seeger et al., 2003). When audiences are highly uncertain about how to react or behave in a crisis, particularly in a deadly and fast-changing crisis, they turn to news media for guidance (Kozman et al., 2021), and this highlights the media's sense-making role in crises. Although legacy news media continues to function as an important source of health and risk information during a crisis, social media also has emerged as an important provider of such information (Ophir, 2018), particularly as it often provides immediate and continuous updates (Wray et al., 2008). Nevertheless, it is important to note that the more people use social media compared to legacy news media during crisis situations, the less factual knowledge they receive about infectious diseases (Kim et al., 2023) and therefore mainstream news media still are considered the most reliable information source during global public health crises.

While news media often adopt a conduit role (Badham, 2019) in transmitting information on behalf of organizations and authorities (Wright, 1960) particularly during ongoing crises, thus giving the impression that organizations and authorities can, to some extent, control news flows (Johansson & Odén, 2018), news media cannot simply be treated as another channel for organizational crisis communicators. Rather, they often act as a political actor (Badham, 2019) and intermediary stakeholder between organizations/authorities and the public. They also have their own source choices, including citizens

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Wigley & Fontenot, 2011) and political actors (Boin et al., 2009), filter information and offer their own interpretations (van der Meer et al., 2017). At times, they can steer the public towards specific choices (e.g., for or against policies of authorities), as discussed in the study by Iannarino et al. (2015) about US public attitudes toward nuclear power development following Japan's 2011 nuclear crisis. Because of this, they are actively shaping how the public constructs an understanding of a crisis, how they interpret the situation and the level of crisis severity, and to whom they attribute crisis responsibility.

Research in crisis framing has outlined an important relation between crisis type, news media framing, and public attitudes (Lim & Seo, 2009), demonstrating that crises considered preventable receive more negative news coverage and lead to more negative public opinion (Iannarino et al., 2015; Lim & Seo, 2009; Turk et al., 2012). The media's crisis framing is thus an important element in shaping responsibility attribution. According to Situational Crisis Communication Theory (Coombs, 2007, 2012), organizations should base their choices of crisis response strategy on the level of responsibility attribution. Crises of external and natural origins such as public health crises would normally be categorized as low in organizational responsibility (Coombs, 2021). However, news media can increase the level of responsibility attribution to organizations through news stories that build a public expectation for organizations' responsibility for crisis management. During the COVID-19 pandemic, for instance, which is considered a public health crisis, many business organizations were expected to contribute to health crisis management, despite the crisis originating beyond national borders (Colleoni et al., 2022). Thus, news media can play a crucial role in how the public interprets a crisis situation and the appropriateness of organizational crisis responses, which heavily influence the public's posture towards an organization.

Substantial research has been conducted since the 1970s on the agenda-setting role of news media (Wanta & Alkazemi, 2017). Agenda-setting theory explains how public perceptions about topics, including public issues and crises, can be influenced by the media, suggesting that news stories published frequently and prominently render them salient and, as a result, audiences consider them important (McCombs & Shaw, 1972). The first level of agenda-setting takes place when media influence the public's perception of what stories or topics matter (Carroll & McCombs, 2003; McCombs, 2004, 2005), while at the second level of agenda-setting, journalists assign attributes to these stories and topics (McCombs et al., 1997). While first-level agenda-setting deals with the visibility of public issues, organizations or political actors, for example, second-level agenda-setting deals with the salience of these attributes on a two-dimensional level: substantive and evaluative (Carroll & McCombs, 2003). According to Weaver (2007), similarities can be seen between second-level agenda-setting and framing, although as processes they are not identical. Both deal with how, for example, people, groups or organizations are represented or treated in the media, and which of these factors are more or less reported in the media.

Crisis literature utilizes the concept of *issue* to describe topics that are shaped or are shaping into specific public opinions and perspectives that can be controversial or contested (Mahon et al., 2018). News topics per se do not constitute an issue unless these topics are contextualized and described in ways that indicate a disagreement over facts, values, and/or policies (Waddock & Mahon, 1991) or a disagreement over procedural or substantive matters related to the distribution of resources or positions (Cobb & Elder, 1983). Yet, during an important crisis such as a global health crisis, news stories commonly include critical topics, ascribe responsibility to political or other social actors, and spread responsibility for containment and support to a broader set of organizations (Thomas et al., 2020). Understanding this dynamic and yet evolving representation of crisis responsibilities and public concerns can be an intricate and time-consuming effort for crisis managers. Nevertheless, crisis managers today have at their fingertips artificial intelligence (AI)-generative tools and other AI-based technologies that can offer insights into the media's agenda-setting role that were not possible a few years ago.

These insights into news media practices and roles are important for crisis communication managers as they help them plan and adjust their communication approaches as crisis events evolve. Within the context of the media's agenda-setting role during crises in general and the COVID-19 pandemic in particular, in the next section, we review key research on the use of AI in collecting and analyzing news articles to derive methodological opportunities to study news media agendas influencing public opinion during crises.

2.2 | Analysing news articles to assess crisis perceptions

Traditionally, public relations professionals track sentiment and trends that may affect their organization's image and reputation by monitoring media coverage for insight into public perceptions of issues, events and organizations. This activity has historically been focused on manual analysis of media content, known as media press clippings (MacNamara, 2005). During recent decades, this activity has expanded its scope to include an organization's entire media landscape which today can be highly digitalized. Because of the large scope of monitoring and assessing media and public agendas across legacy and digital media outlets, automated content analysis techniques have emerged and gained traction in journalism and media studies (e.g., Boumans & Trilling, 2018; Burggraaff & Trilling, 2020), including research in crisis communication (e.g., De Bruyne et al., 2024; van der Meer, 2016).

Agenda-setting studies tend to limit media analyses to just a handful of topics or issues hand-picked by researchers (Kim et al., 2014), which are time-consuming in their analyses of small-scale data sets and manual categorization into positive and negative frames. Automated approaches (including supervised and unsupervised techniques) have addressed this long-standing limitation in media research, with agenda-setting research transitioning to larger data sets and computational methods. There is growing interest in the use of computational tools for efficient and large-scale analysis of media agendas

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Korenčić et al., 2015). Larger data sets and in-depth text mining of online news can in fact yield more meaningful results, offering an effective alternative to traditional agenda-setting research (Kim et al., 2014). Notably, researchers have highlighted the significant advancements made in natural language recognition and computation (e.g., van der Meer et al., 2016, 2017; Page et al., 2023; Zhang et al., 2023), emphasizing their relevance in studying human communication patterns. The rise of AI technologies has further expanded the possibilities for analysing textual data, particularly through inductive, yet quantifiable methods (Blair et al., 2020). Among these, it is noteworthy to mention NLP as a fundamental area of study in the realm of AI and computer science, encompassing theories and techniques that facilitate efficient communication in natural language between humans and computers. NLP has been utilized specifically to analyse large data sets of mediated communications, for instance social media posts (e.g., Blair et al., 2020), as well as news media articles (e.g., DiMaggio et al., 2013).

One widely used approach to extract information from large textual material utilizing NLP is topic modelling. Topic modelling is one of the most used unsupervised approaches that essentially involves statistical methods to identify patterns of word co-occurrence and to group words together into topics that represent coherent themes within the corpus (Albalawi et al., 2020; Vayansky & Kumar, 2020). One significant benefit is that, unlike supervised methods, topic inference is based on the provided collection without prior knowledge. This approach helps sidestep what is referred to as the 'streetlight syndrome' (Maier et al., 2018). In recent years, various methodologies for topic modelling have emerged, among which Latent Dirichlet Allocation stands out as the most recognized and commonly utilized algorithm (Jockers & Thalken, 2020). Additionally, there are lesser-known methods such as latent semantic analysis (Landauer et al., 1998), structural topic modeling (Lindstedt, 2019), non-negative matrix factorization (Wei et al., 2003), correlation explanation (Gallagher et al., 2017), Top2Vec (Angelov, 2020) and BERTopic (Grootendorst, 2021), among others, which can provide superior performances under specific conditions. According to Jacobs and Tschötschel (2019), topic modelling helps to overcome the practical barriers of making sense of how topics in news media evolve and contributes to methodological rigour and systematic meaning-making related studies.

During the COVID-19 pandemic, NLP methods have been employed across various studies to analyse public sentiment and attitudes on platforms like Twitter/X (SV & Ittamalla, 2021), examine topic distributions in news articles and commentaries (Wang & Mengoni, 2020; Wu, 2021) and assess the prevalence of specific COVID-19 themes in media coverage over time (Lu et al., 2023; Tomasi et al., 2023). Furthermore, crisis research has utilized topic modelling to explore citizen concerns expressed through social media during a public health crisis (Deng et al., 2020), evaluate political bias in the news coverage (Xu, 2023), analyse social media responses to events like terrorist attacks (Toivanen et al., 2020) and investigate media framing during crises such as the European refugee crisis (Heidenreich et al., 2019). Much of the current NLP studies on news articles have focused on first-level agenda-setting and were localized to large language groups, with many studies based on English-language text corpora. However, understanding the second-level agenda-setting role of news media is vital to plan effective crisis responses. Furthermore, the use of NLP methods to capture both agenda-setting levels in minor language groups is less known. Based on the extant literature and research gaps, we are thus interested to find out:

RQ1. How does NLP contribute to a media analysis of a lengthy pandemic crisis situation?

RQ1a. How does NLP identify the main crisis topics reported (first-level agenda-setting) and treated (second-level agenda-setting) by the news media of specific countries (in this case, Finland and Sweden) throughout a 2-year pandemic (in this case, COVID-19)?

RQ1b. How does NLP identify differences in the way the news media of specific countries (in this case, Finland and Sweden) reported (first-level agenda-setting) and treated (second-level agenda-setting) these crisis topics throughout a 2-year pandemic?

3 | RESEARCH APPROACH

To answer this study's research questions, we utilized news articles about the COVID-19 pandemic in Finnish and Swedish news media over a 2-year period. The news articles were analysed through topic modelling, specifically Top2Vec method (Angelov, 2020). Top2vec is a topic model that works by constructing a shared embedding for a set of documents and words (Le & Mikolov, 2014; Mikolov et al., 2013) such that proximity in the embedding indicates similarity for both documents and words (Angelov, 2020). In topic modelling literature, 'topics' represent sets of words that semantically group similar documents, such as news articles. In this study, topics were constructed as issues related to the COVID-19 pandemic. Crisis literature utilizes the concept of 'issues' to describe critical topics that are shaped or are shaping into specific opinions and perspectives that can be controversial or contested (Mahon et al., 2018). News articles were the unit of analysis, where multiple topics and related co-occurrence of words were then extracted and grouped.

3.1 | Data collection procedure

The data set is part of a larger research project investigating COVID-19 communications across multiple countries (Colleoni et al., 2022). Data was acquired through the company Retriever, which is the largest Nordic digital news archive. Data includes all online news articles on COVID-19 published in Finland and Sweden, during the period February 2020 and June 2022, published in the following news outlets: For Finland, *Iltalehti*, *Ilta-Sanomat*, *Yle*, *Hufvudstadsbladet* and *Svenska Yle*. The latter two outlets were included because Swedish is an official language in Finland, spoken by a minor and yet relatively important group of Swedish-speaking Finns (Statistics Finland, 2023). For Sweden, data was collected from the

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

following news outlets: *Svenska Dagbladet*, *Aftonbladet*, *Expressen*, *SVT Nyheter* and *Dagens Nyheter*. These newspapers were selected based on their size of circulation and their representation of different voices, reflecting the spectrum of media coverage in Finland (Reunanen et al., 2023) and Sweden (Westlund, 2023). The searched keywords for Finland and Sweden included the following:

> 'sars-cov-2' OR 'covid-19' OR ((koronaviru* OR coronaviru* OR covid*) AND (pandemia* OR epidemi* OR rokotu* OR rokot* OR karanteen* OR eristäy* OR eristä* OR eristy* OR tartuttami* OR tartun* OR pandemin* OR epidemin* OR vaccination* OR vaccin* OR karantän* OR isolering* OR smittning* OR smitta*))

Overall the analysis of (1) Finnish news articles (consisting of 32,772 online news articles in Finnish language), (2) 11,487 articles in Finnish-Swedish language and (3) 50,583 news articles from Swedish outlets.

3.2 | Data preprocessing and analysis

The raw data was preprocessed for the downstream analysis separately in Finnish and Swedish languages. We used Python 3.6.15 in all preprocessing and analysis steps. The raw data concerning the Finnish language news texts were preprocessed using Turku Neural Parser (Kanerva et al., 2018). The parser pipeline was used for tokenization, lemmatization (Kanerva et al., 2021) and for morphological tagging. In tokenization, a long string (e.g., an article) is split into multiple substrings, typically words and punctuation marks. Lemmatization refers to transforming a word into its dictionary form, that is, lemma (see, e.g., Jurafsky & Martin, 2008). Swedish language articles were preprocessed similarly using the Trankit toolkit (Van Nguyen et al., 2021).

We used lemmatized articles only for collecting the named-entities in their base form. For analysing the topics and named-entity recognition (NER), we used the simply tokenized documents as an input. Key social actors and organizations present in the articles were identified using the NER software Stanza (ver. 1.4.2) (Qi et al., 2020). Top2vec (Angelov, 2020) was employed for topic modelling. The topics are found by finding dense clusters of similar news articles and computing a topic vector for each cluster. Finally, the words characterizing each topic can be found by selecting the closest word vectors in the embeddings space (Angelov, 2020), as the similarity measure we used the cosine similarity.

We used the Gensim Python library (Řehůřek & Sojka, 2010) to simultaneously train word and document vectors following the word2vec (Mikolov et al., 2013) and doc2vec (Le & Mikolov, 2014) algorithms and the collected and tokenized Finnish and Swedish article data and the Python implementation of HDBSCAN 0.8.29 (Campello et al., 2013) for finding the document clusters. Top2vec generates a variable number of topics based on the data and model parameters, due to how HDBSCAN operates. In the clustering phase, we set the minimum cluster size to 20 to avoid picking up noisy or very specialized topics. Initially, we identified 138 topics for the Finnish data set, 53 for the Finnish-Swedish data set, and 181 for the Swedish data set. We then reduced the number of topics to a manageable size by iteratively merging the smallest topic with the most similar topic, as measured by Euclidean distance, following Angelov's (2020) recommendations. The final number of topics was selected based on the UMass coherence measure (Mimno et al., 2011) within the range of 10–20.

Next, topics were associated to reflect the first level of agenda-setting, which pertains to the overall issue, event or actor discussed regarding COVID-19. Following van der Meer (2016) and Leydesdorff and Hellsten (2006), we treated the co-occurrences of words associated with these topics, also referred to as word associations (Murphy, 1983; Rogers et al., 2021), as discursive construction of crisis meanings describing topic attributes, thus representing the second level of agenda-setting. While word co-occurrence has been used in studying framing, we followed Weaver's (2007) perspective and considered framing as a part of agenda-setting that operates as a 'second-level' or secondary effect. Thus word co-occurrences served to identify the attributes associated with the main topics, the frequency allowed to detect what news media wanted people to remember the most about the topic. For example, the co-occurrence of words such as *kindergartens* and *schools* with *virus exposure* indicates that the Finnish news media specifically wanted people to think about virus exposure in the context of kindergartens and schools. Next, we looked at second-level agenda-setting by analysing the most common words of every crisis topic. By looking at the frequency of words in each topic, it was possible to see the most central attributes of each topic (Kim et al., 2012). Again, using the same example of virus exposure in schools and kindergartens, the presence of words such as *exposure*, *being exposed to*, *hazard*, *chain of contagion*, and so forth, indicate that the news media prompted people to think about places such as schools and kindergartens in terms of 'COVID-19 risk' and 'contagion hazard'. To simplify the presentation of the results, we limited the analysis to the 15 most frequently used words to see which perspectives the media considered important in each topic.

4 | RESULTS

In this section, we present the results of the topic modelling analysis focusing first on identifying the dominant crisis topics reported and treated by Finnish and Swedish news media, which relates to the first level and second level of the media's agenda-setting role (addressing RQ1a), and second on identifying differences in media reporting and treatment between the two countries (Finland and Sweden) (addressing RQ1b). Collectively, the description and discussion of these results serves to demonstrate how NLP contributes to media analysis of a pandemic crisis situation (addressing RQ1).

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.1 | First-level agenda-setting: What news media talked about

4.1.1 | Finland

In the Finnish-language media, a total of 15 dominant crisis topics were identified through the analysis of the Finnish data (see Figure 1). The most reported crisis topic was 'business restrictions' ($n = 5163$ articles), which was covered by all three Finnish news media (Yle, Iltalehti and Ilta-Sanomat). This topic dealt with the government's decisions to place restrictions on business activities. 'Exposure of the virus in educational institutions (e.g., schools)' was the second most covered topic ($n = 3615$), mostly due to Yle's extensive coverage. The third most covered crisis topic ($n = 3517$ articles) was 'global economic impact'. The topic of 'culture and entertainment' ($n = 2508$) and the 'Finnish ice hockey league' ($n = 2460$) were covered more than 'vaccine and vaccinations' related topics ($n = 2052$). Other topics listed in Figure 1 (below) attracted less coverage but still somehow gained quantitatively equal coverage from each Finnish news outlet.

Figure 2 shows topic salience along a timeline (February 2020 to June 2022). The COVID-19 pandemic caused a large spike in news articles in both Finnish and Finnish-Swedish news media in the spring of March–May 2020. The emergence and rapid spread of the virus led to a high level of public demand for information about the pandemic. Constantly changing and extensive restriction measures increased news coverage and discussion of various pandemic-related topics. However, throughout the pandemic Finnish news coverage focused mostly on policies and business restrictions as well as measures in schools and workplaces.

In the Finnish-Swedish news media in Finland (see Figure 3), 20 different crisis topics were covered. Two crisis themes stood out from the data: 'COVID-19 and foreign political elite' ($n = 1561$ articles), which included news linking the pandemic with political decision-makers in different countries, and 'domestic policies' ($n = 1316$ articles), which related to Finnish political decision-making and various restrictions in the population. The third most reported crisis topic was 'experts and health care organizations' ($n = 815$ articles), which included news related to various healthcare professionals and organizations. Finnish-Swedish news media covered this topic equally. Both news outlets covered the topic of 'vaccines and vaccinations' ($n = 815$ articles), which contained themes and actors (e.g., heads of vaccine manufacturers) related to the COVID-19 vaccine, and 'COVID-19 and economics' ($n = 737$ articles) regularly, which included various perspectives on the economic impact of the pandemic.

Figure 4 shows the Finnish-Swedish media's topic salience along a timeline (February 2020 to June 2022). They focused mostly on reporting the political decision-making of authorities in other countries as well as the policies and business restrictions in Finland. Changes in the amount of coverage were linked to periods in which virus variants emerged in the population. The Finnish-Swedish media also highlighted more healthcare experts and organizations which played a prominent role in both Finnish and Finnish-Swedish news media. In the middle of the pandemic, vaccinations and the vaccines became a significant topic of discussion in the Finnish-Swedish media.

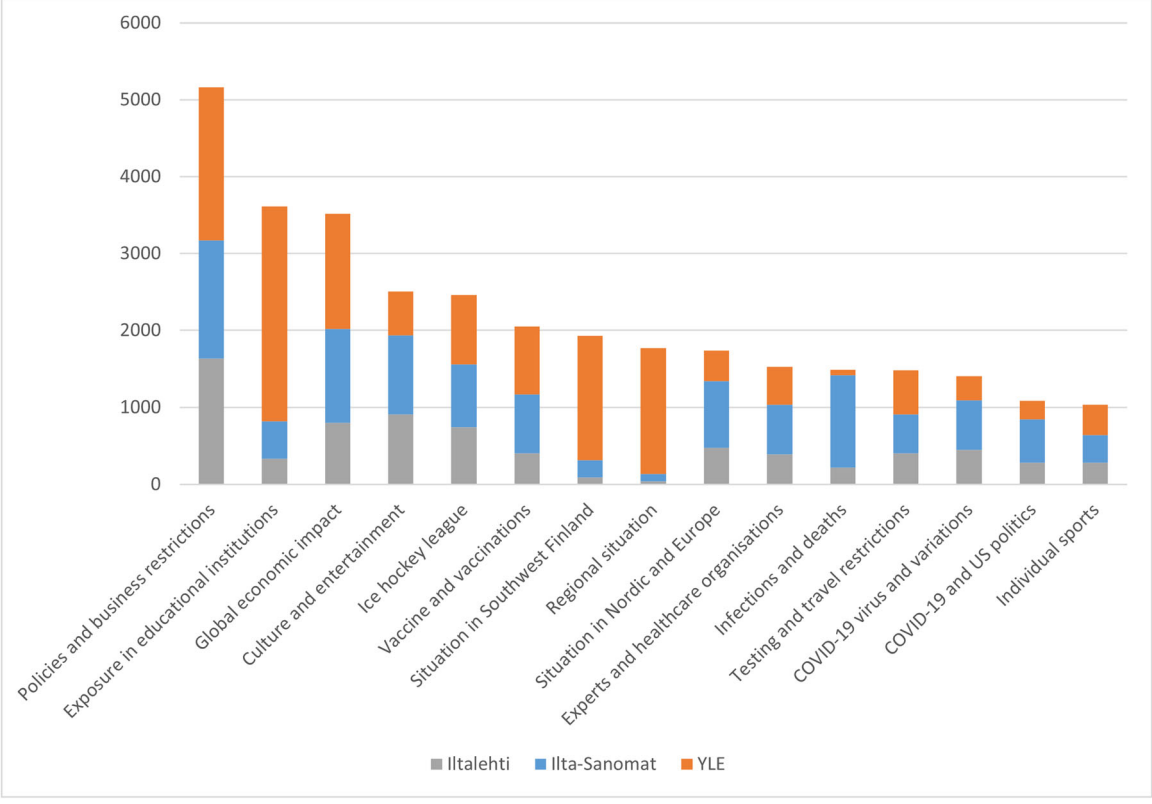

FIGURE 1 Crisis topics in Finnish news media. *Source*: Own elaboration.

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

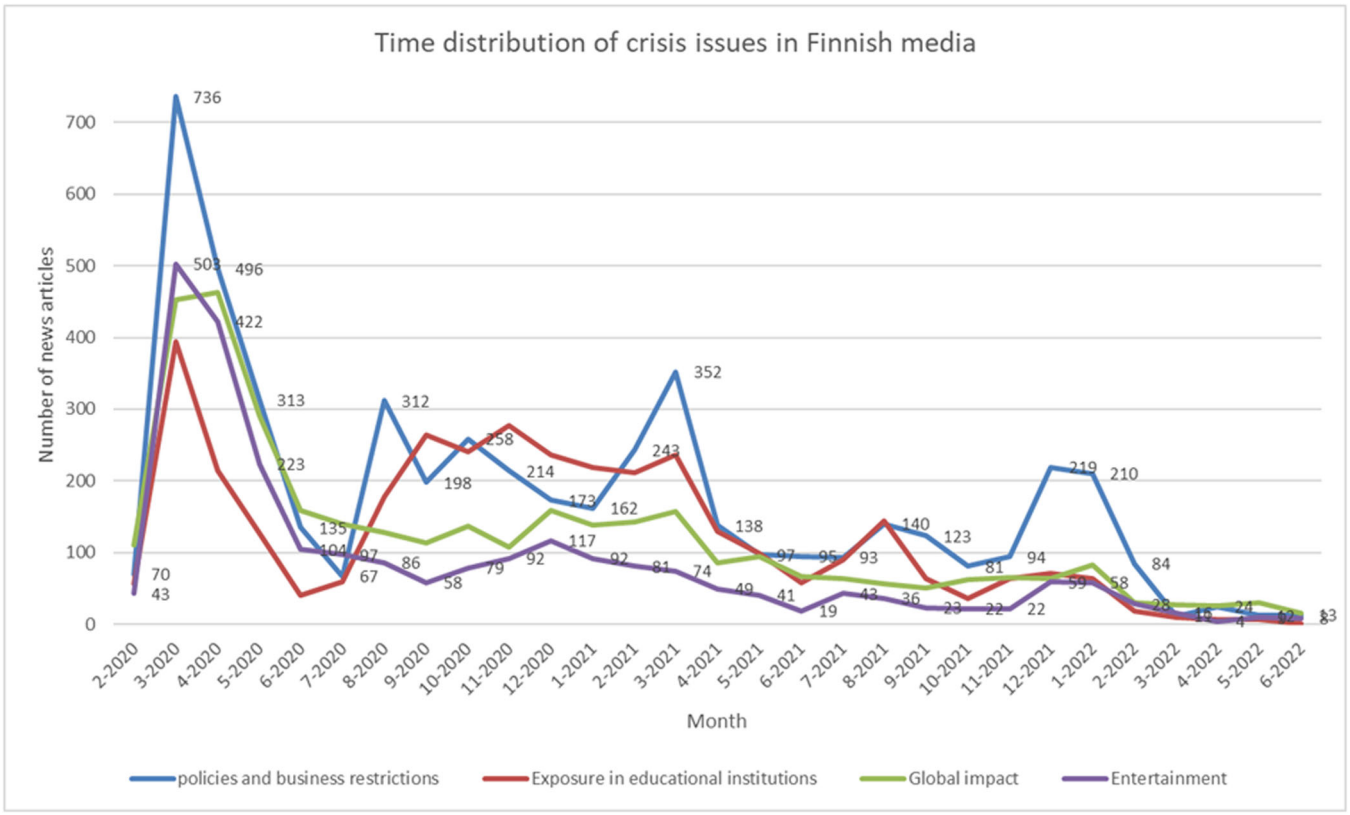

FIGURE 2 Main topic distribution in Finnish news media. *Source*: Own elaboration.

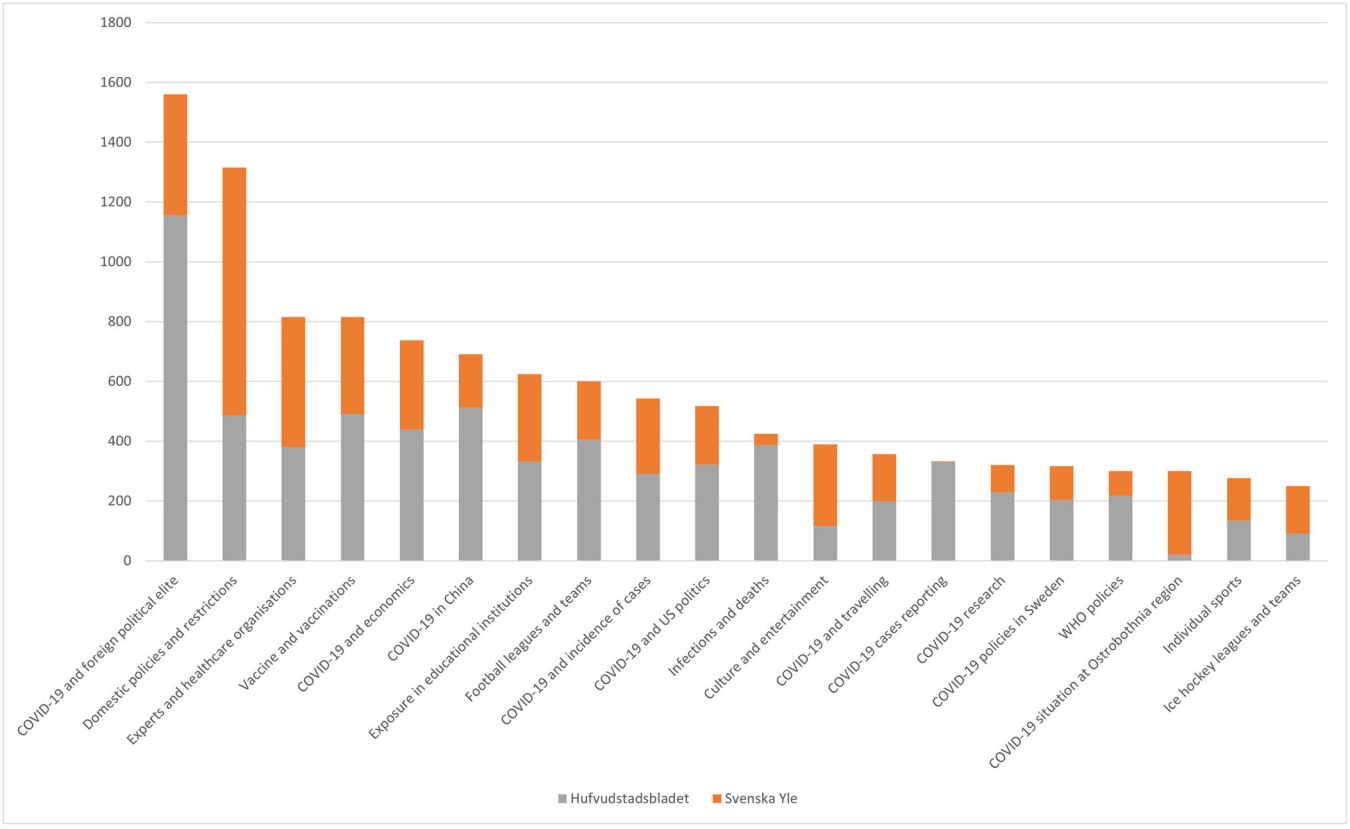

FIGURE 3 Crisis topics in Finnish-Swedish news media. *Source*: Own elaboration.

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

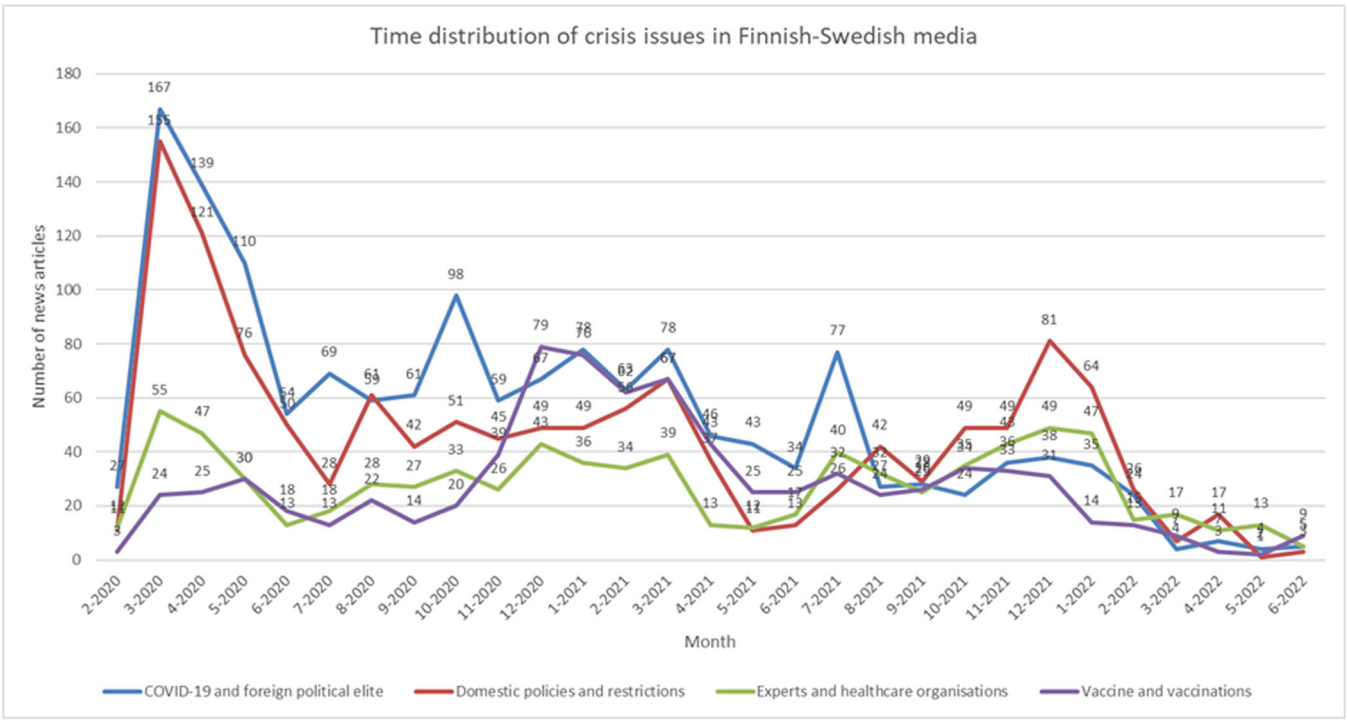

FIGURE 4 Main topic distribution in Finnish-Swedish news media. *Source*: Own elaboration.

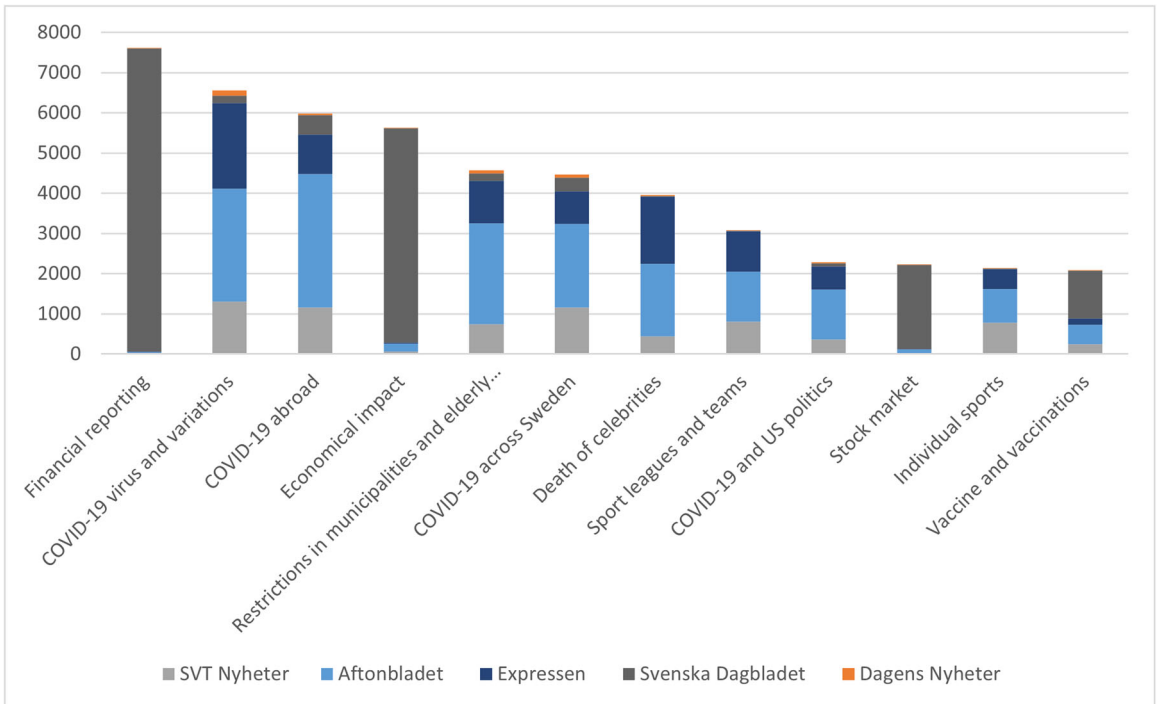

FIGURE 5 Crisis topics in Swedish news media. *Source:* Own elaboration.

4.1.2 | Sweden

In Sweden, 12 crisis topics were prominently covered by Swedish media (see Figure 5). The total number of news articles in the Swedish-language material was high ($N = 50{,}583$), and four major topics stood out in particular: 'financial reporting' ($n = 7605$), 'COVID-19 virus and variations' ($n = 6558$), 'COVID-19 abroad' ($n = 5977$) and 'economical impact' ($n = 5615$). The 'financial reporting' topic included content related to publicly listed companies. Overall, the results show that the coverage of COVID-19 topics was

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

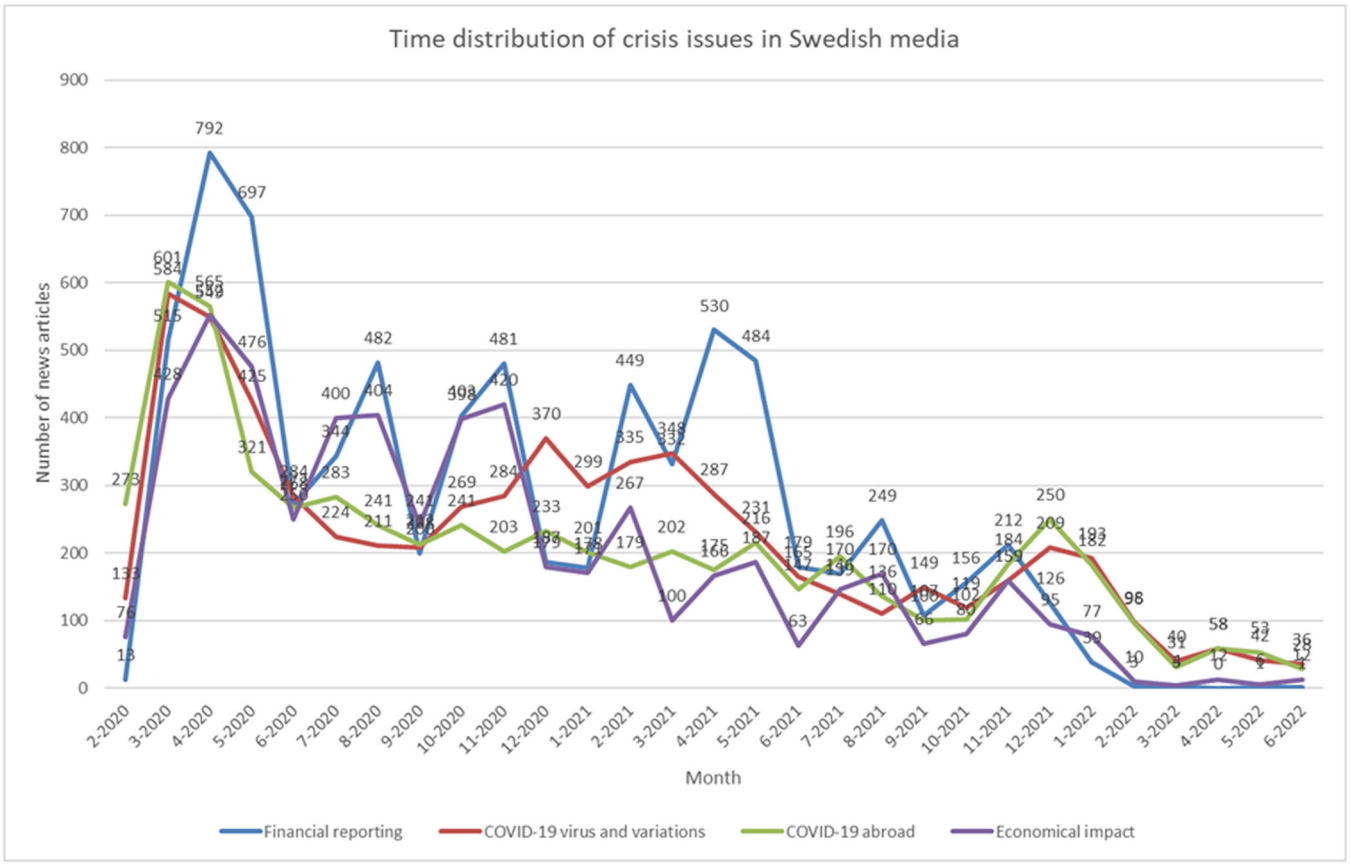

FIGURE 6 Main topic distribution in Swedish news media. *Source*: Own elaboration.

strongly divided between two news media. *Aftonbladet* reported a lot about the virus, the domestic situation, restrictions, the situation in foreign countries and topics related to sports. *Svenska Dagbladet*, on the other hand, focused on topics related to the economy and vaccinations.

Figure 6 shows the Swedish media's topic salience along a timeline (February 2020 to June 2022). News media reports about COVID-19 focused mostly on the economic effects throughout the pandemic. This can be seen in quarterly financial reporting, but also in a steady amount of news about broader economic effects. The emergence of virus variants in the population was evident in news spikes. The 'virus and its variants' topic became significant news for a long time, especially with the development and arrival of vaccines. The media in Sweden also regularly reported on the pandemic situation abroad.

4.2 | Second-level agenda-setting: *How* news media talked about COVID-19 related topics

Our topic modelling analysis identified 15 dominant words in news articles that serve to define and frame topics with specific attributes, thus offering insights related to second-level agenda-setting (McCombs et al., 1997).

4.2.1 | Finland

Finnish news media (see Table 1) associated words such as customer *occupancy limit*, *infectious disease law* and *hybrid strategy* to the topic of *domestic policies* and *restrictions*. News media also frequently associated social actors such as the former Finnish prime minister Sanna Marin, the Minister of Social Affairs and Health and other local actors to this topic. The fact that these social actors received extensive media attention indicates the news media's role in ascribing topic responsibility to certain individuals or institutions (SV & Ittamalla, 2021; Thomas et al., 2020). Media covered the COVID-19 pandemic in educational institutions mostly by reporting the amount of virus exposure in schools and kindergartens. Finnish language news media were particularly interested in monitoring the exposure of children and young people to the virus and associated words like *exposure*, *chain of transmission* and *kindergarten* and *schools* to this topic. In this topic, a certain infectious disease doctor (*Markku Broas*), represented as an expert commentator, was commonly a co-occurring word.

The economic impact topic was often associated with words related to geopolitical contexts, such as the European Union and China. Attributes associated with this topic were: *economy*, *Eurozone*, *European Central Bank*, *China*, *gross national product* and *world economy*. These attributes indicate that the European perspective was

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Attributes of topics in Finnish news media.

Topic	Attributes (15)—Finnish (and English equivalent)
Domestic policies and restrictions	Asiakaspaikkarajoitus [customer occupacy limit], tartuntatautilain [infectious disease law], vappupiknikiin [first of May picnic], hybridistrategian [hybrid strategy], eduskunnassa, eduskunnan, eduskunnalle [parliament], aluehallintovirastot, avin [Regional State Administrative Agency], marin [Sanna Marin], stm [The Ministry of Social Affairs and Health], ravitsemisliikkeiden [nutrition shop], harrastajateatteriin [amateur theater], perustuslakivaliokunta, perustuslakivaliokunnan [constitutional committee]
Exposure in educational institutions	Tavoitettu [reached], altistuneet, altistuneita, altistuneisiin, altistuneiden, altistumisia, altistunutta [exposed], asetettu [placed], tartuntaketjuun [chain of transmission], päiväkodissa, päiväkodin [kindergarten], koululla, koulussa, [school], broaksen [Broas'], lievissäkin [even in mild]
Global economic impact	Talouden, taloutta, talous [economy], euroalueen [euro zone], ekp [European Central Bank, ECB], sodan [war], keskuspankki [central bank], kiinaa, kiina [China], lagarde, kriisin [crisis], bruttokansantuote [gross national product], valtioiden [state], Xi, maailmantalouden [world economy]
Culture and entertainment	Peuhu, prinssi [prince], kuningatar, kuningattaren, [queen], herttuatar [duchess], perhe [family], laulaja [singer], pari, pariskunta [couple], Elisabet, näyttelijä [actress], minulle, minulla, minä, [me] Meghan
Ice hockey league	Liigan, liiga, liigassa, sm-liigan [league], pelata [play], ottelu, ottelun, ottelua [match], pelaajaa [player], joukkue [team], pelaajat, pelaajien [players], joukkueen, joukkueiden [team], seuran [club]
Vaccine and vaccinations	Rokote, rokotteelle, rokotteella, rokotteesta, rokotetta, rokotteen [vaccine], pfizerin [Pfizer's], modernan [Moderna's], astrazenecan [Astra-Zeneca's], biontechin [Biontech's], myyntilupaa, myyntilupa, myyntiluvan [marketing authorization], lääkeyhtiö [pharmaceutical company], ema [European Medicines Agency]
Situation in Southwest Finland	Varsinais-suomessa, varsinais-suomen [Southwest Finland], alueellinen [regional], nyrkki, koronanyrkki [corona fist], pandemiaohjausryhmä [pandemic steering group], etelä-pohjanmaalla [South Ostrobothnia], sairaanhoitopiirin, sairaanhoitopiirissä [medical district], edelleen [still]], koronakoordinaatioryhmä [corona coordination group], kiihtymisvaiheeseen, kiihtymisvaiheessa [acceleration phase], maakunnassa [province], johtajaylilääkäri [medical director]
Regional situation	Pohjois-pohjanmaa, pohjois-pohjanmaalla [North Ostrobothnia], katsele [watch], kainuusta, kainuun, kainuussa [Kainuu], tapahtuu [happening], pirkanmaalla [Pirkanmaa], lähelläsi [near], Koukkari, mitä [what], uutisvahti [news watch], tästä [here], liittyvät [related], jutusta [story]
Nordic and European situation	Maan, maassa, [country], Löfven, Conte, valtionepidemiologi [state epidemiologist], ruotsissa [Sweden], Tegnell, Tegnellin, reutersin [Reuters], italiassa [Italy], ranskassa [France], victoriassa [Victoria], osavaltion [state], guardianin [Guardian], Hancock
Experts and healthcare organizations	Voipio-pulkki, Voipio-Pulkin, Liisa-Maria, strategiajohtaja [strategy director], sairaalahoidon [hospital care], tehohoidon [intensive care], Mäkijärvi, Reinikainen, Reinikaisen, kuormitus [load], tapausmäärät [number of cases], husin, hus-alueella [Hospital District of Helsinki and Uusimaa (HUS)], Helve, Bendel
Infections and deaths	Raportoitu [reported], eilisestä [from yesterday], kuolemia, kuolemantapauksia, kuolleiden [deaths], luku [number], sairaalahoidossa [hospitalized], raportoitiin [reported], tartuntoja, tartuntojen, [infections], tehohoidossa [intensive care], pyöristettyjä [rounded], varmistettuja [confirmed], edellisten [previous], tapauksia [cases]
Testing and travel restrictions	Matkustajien, matkustajia, matkustajat, matkustajille, matkustajaa, matkustajan, matkustaja [passengers], laivalla, laivoilla [on board], siljan [Silja], satamassa [at the port], Tallink, Nöjd, lentoasemalla [at the airport], Boijer-Svahnström
COVID-19 virus and variations	Tutkimuksessa, tutkimus, tutkimuksen [research], mutaatio, mutaatioita, mutaatiota, [mutation], tutkijat, tutkijoiden [researchers], variantti [variant], vasta-aineita, vasta-aineet [antibodies], infektion [infection], elimistössä [in the body], muunnos [transformation], immuniteettia [immunity]
COVID-19 and US politics	Trump, trumpin, Trumpia [Trump], Donald, Biden, Bidenin, Joe [Biden], demokraattien [Democratic party], valkoisen, valkoisessa [White House], republikaanien [Republican party], yhdysvaltain [United States], presidentti, presidentin [president], kongressin [Congress]
Individual sports	Kisojen, kisoja, kisat, kisoihin, kilpailut, kisan, kisaa [games], olympialaisten, olympialaisiin, olympialaisissa, olympialaiset, [Olympics], maailmancupin [world cup], tokion [Tokio], urheilijat [athletes], osakilpailua [competition]

Source: Own elaboration.

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

therefore particularly important to the news media, but also the economy of the country of origin of the disease and the general development of the global economy.

Culture and entertainment topics were mostly associated with the royal family, celebrities and singers. In using these attributes, the media used human interest-style reporting in its COVID-19 coverage. Ice hockey is a strong cultural phenomena in Finland and this topic was extensively covered in articles using words associated with teams, matches and players. Additionally, words related to cancellation of matches show that the media linked the impact of the pandemic to sport entertainment.

Vaccines and vaccinations were mainly associated with attributes related to medical companies and sales permits. In this topic, various vaccine manufacturers dominated the news coverage. The COVID-19 situation in Southwest Finland was covered from an operational perspective with a focus on the local medical district. Articles contained attributes referring to *steering group*, *corona coordination group*, *corona fist* and *acceleration phase*, which indicate that news media actively followed regional treatment of the corona situation. The topic of experts and healthcare organizations was covered a lot, concentrating on certain experts and hospital districts. Attributes listed in this topic are to a significant extent names of doctors who were cited in news articles as expert commentators.

Compared to the Finnish-language news articles, the Swedish-language articles published in Finnish media utilized slightly different attributes to discuss the crisis topics. The dominant topic of 'COVID-19 and foreign political elite' (see Table 2) that focused mostly on the key political decision-makers of various countries contained words associated mostly to political actors, particularly prime ministers in various countries (e.g., Boris, Johnson, Rutte, Yoshihide, Castex), indicating media attention to decision-makers' roles and responsibilities in pandemic management. The second dominant topic was 'pandemic policies and restrictions', which included COVID-19 passports or certificates. The attributes here were restrictions, a spreading phase, corona passport and a constitutional committee. The main political actor associated with this topic was Prime Minister Sanna Marin. This topic contained attributes related to educational institutions as well. The topic of 'experts and healthcare organizations' also received extensive coverage, as it did in Finnish news media coverage where individual pandemic-related experts received most of the news attention. Attributes here are highly related to the same experts as in Finnish news media, who were expert commentators. The topic of 'vaccines and vaccination' received extensive attention, as it did in Finnish news media, by referring predominantly to vaccine companies. The attributes (e.g., Astra-Zeneca, Pfizer, Moderna, Biontech and European Medicines Agency) were related to anticipation of vaccines and sales permits. The dominant 'COVID-19 and economics' topic was given a broader perspective than Finnish news media. The list of attributes included words such as *Eurozone*, *growth*, *financial crisis*, *global*, *investment*, *employee turnover* and *GDP*. The 'COVID-19 in China' topic was formed around words such as *Wuhan*, *Hubei* and *Hong Kong*, which suggests that the coverage was focused on the supposed source of the virus. 'Exposure in educational institutions and schools' was another dominant topic in Finnish-Swedish news media, with the attribute selection and variation indicating more of a focus on upper secondary schools.

4.2.2 | Sweden

The most covered topic in Swedish news media was 'financial reporting', and consisted of attributes related to publicly listed companies, such as Empir, Bultens and Gränges, which are companies that were highly influential or highly affected by the pandemic (see Table 3). In the topic of 'COVID-19 virus and variations', the associated words were *virus immunity*, *mutations* and certain medical experts such as various virologists' names. Swedish news media also substantially dedicated news space to cover the COVID-19 situation abroad. Topic modelling also revealed the source of such information, which were various news agencies, as many words associated with this topic relate to AFP, Reuters and AP press agencies. Restrictions in municipalities and elderly care facilities were covered much more in Swedish media than in Finnish and Finnish-Swedish media. Attributes in this topic (e.g., *elderly care*, *nursing home*, *home care* as well as *municipalities)* indicated that news media wanted to show how the pandemic affected vulnerable groups of citizens. Tabloid-like news reports are evidenced in the news coverage of topics such as the 'death of celebrities'. The attributes (e.g., *beloved*, *love*, *sorrow* and *loss*) in this topic contained very emotional emphases and referred to two TV personalities who died of COVID-19 disease. The 'COVID-19 and US politics' topic, which was also covered by Finnish and Finnish-Swedish news media, was covered similarly, emphasizing the role of the two presidents (Trump and Biden) as key decision-makers in the pandemic. Additionally, the topic of 'vaccines and vaccinations' follows the same pattern as Finnish and Finnish-Swedish news media. The attributes here were *Astra-Zeneca*, *Pfizer*, *Moderna*, *Biontech* and *European Medicines Agency*, which indicate that in Sweden news media also anticipated the vaccine and the sales permit.

4.3 | Differences across Finland and Sweden

The analysis of topics (first-level agenda-setting) and word associations (second-level agenda-setting) covered by the different news media shows that NLP analysis via topic modelling is suitable to identify trends and patterns of news coverage of a public health crisis and can outline important differences in media reporting and treatment of the same crisis situation across two countries (i.e., Finland and Sweden) and their media outlets, thus addressing RQ1b. Indeed, comparing the agenda-setting role of news media across Finland and Sweden contributes to the limited extant research investigating multicountry agenda-setting contexts (McCombs, 2005). While some similarities are visible in the topics covered by the Finnish and Swedish media, other substantial differences can be traced through this method, showing that news media still tend to localize discussions and interpretations of crisis situations based on local and

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Attributes of topics in Finnish-Swedish news media.

Topic	Attributes (15)—Finnish-Swedish (and English equivalent)
COVID-19 and foreign political elite	Premiärminister, premiärministern, [prime minister] Boris, Yoshihide, Castex, Rutte, Conte, Sánchez, Downing, hälsominister [health minister], Johnson, japanska [japanese], Gladys, parlamentet [parliament], Dock
Domestic policies and restrictions	Begränsningar [restrictions], Marin [Sanna Marin], spridningsfasen [spreading phase], coronapasset [corona passport], grundlagsutskottet [constitutional committee], covidintyget [covid certificate], restaurangerna [restaurants], koordineringsgruppen [coordination group], hobbyverksamhet [hobbies], regionförvaltningsverken [regional administrative bodies], hybridstrategi [hybrid strategy], samlingspartiets [coalition parties], nyland [Uusimaa], läroanstalter [educational institutions], huvudstadsregionen [capital region]
Experts and healthcare organizations	Voipio-pulkki, Puumalainen, Liisa-Maria, Taneli, Rintala, Pietilä, Åucs, smittfallen [infections], Salminen, egentliga [actual], Helve, överläkare [chief physician], strategidirektör [director of strategy], Katz, universitetscentralsjukhus [university central hospital]
Vaccine and vaccinations	Vaccinet, vaccin [vaccine], astra, zeneca, zenecas [Astra-Zeneca], pfizer, pfizers [Pfizer], läkemedelsmyndighet, läkemedelsmyndigheten [pharmaceutical authority], ema [European Medicines Agency], modernas [Moderna], biontechs [Biontech], läkemedelsbolaget [pharmaceutical company], fda, försäljningstillstånd [marketing authorization], Nohynek
COVID-19 and economics	Ekonomin [economy], eurozonen [Eurozone], kvartal, kvartalet [quarter], tillväxt, tillväxten [growth], ekonomiska [economic], finanskrisen [financial crisis], globala [global], investeringar [investment], omsättningen [turnover], sysselsättningen [employment], bnp [GDP], inköpschefsindex [purchasing managers' index], krisen [crisis]
COVID-19 in China	Wuhan, kina [China], kinesiska [Chinese], hubeiprovinsen, hubei, provinsen [Hubei province], Yokohama, livsmedelsministern [food minister], hongkong [Hong Kong], utbrottet [outbreak], kryssningsfartyget [cruise ship], who [WHO], virologiska [virological], Peking, djur [animal]
Exposure in educational institutions	Exponerats, exponerade [exposed], sipoonjoen [River Sipoo], koulu, skola, skolans, skolan [school], försatts [placed], kontaktats [contacted], utsatts [exposed], lukio [high school], elev, elever [student(s)], hemkarantän [home quarantine], beväringen [conscription]
Football leagues and teams	Klubben [club], spelarna, spelare [players], premier, matchen, match, matcherna [match], ligan, league [league], Arsenal, Tottenham, spela [play], Brighton, champions, laget [team]
COVID-19 and incidence of cases	Tvåveckorsperioden, tvåveckorsperiod [two-week period], nu [now], bokförts [recorded], nya [new], att [that], incidensen [incidence], coronasaldo [corona stats], vårdas [maintained], registrerats [registered], Loimaa, coronasiffra [corona number], enligt (according to], det [the], fall [case]
COVID-19 and US politics	Trump, Donald, Trumps [Trump], Biden, Bidens, Joe [Biden] presidenten [president], usa:s [United States], demokraterna, demokraternas [Democratic party], vita, husets, huset [White House], amerikaner [american], Fox
Infections and deaths	Värst [worst], döda [dead], avlidit, avlidna [deceased], drabbade [affected], landet [country], passerat [passed], samtidigt [simultaneously], dödstalet [death toll], människor [people], brasilien [Brazil], antalet [number], kina (China], dygnssiffran [daily figure], indien [India]
Culture and entertainment	Roligt [funny], musik, musiker, musiken [music], låtar [songs], jag, mig [me], Westö, gemenskap [community], känns [feels], Merete, Saine, Alva, skrattar [laughs]
COVID-19 and travelling	Schengenländer [Schengen countries], resenärer, resenärerna [travelers], Skopje, schengenområdet [Schengen area], inresa, inresor [entry], inreserestriktionerna [entry restrictions], passagerare, passagerarna [passangers], eu- [EU], anländer [arriving], nordmakedonien [North Macedonia], reserestriktionerna [travel restrictions], flyg [flight]
COVID-19 and cases reporting	Fördröjning [delay], laboratorieresultat [laboratory results], alltså [that is], institutets [institute], anmälarens [notifier], incidensen [indidence], coronasiffrorna [corona figures], sjukvårdsdistrikten [healthcare district], dataöverföringen [data transfer], rapporteras [reported], enligt [according to], anmälaren [notifier], Päijät-Häme, registret [register], meddelar [notifies]
COVID-19 virus and variations	Studien, studie, studier [study], forskarna, forskare [researchers], antikroppar, antikropparna [antibodies], infektion [infection], immunitet, immuniteten [immunity], neutraliserande [neutralizing], återinfektion [reinfection], karolinska [Karolinska Institutet], tidskriften [journal], forskningen [research]
COVID-19 policies in Sweden	Folkhälsomyndigheten, folkhälsomyndighetens [The Public Health Agency], Tegnell, Hallengren, sverige [Sweden], Carlson, statsepidemiolog [state epidemiologist], socialstyrelsen [social welfare board], Löfven, socialminister [Social minister], Stockholm, Wallensten, Byfors, Anders, Tegmark

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Topic	Attributes (15)—Finnish-Swedish (and English equivalent)
WHO and COVAX policies	Fattiga, fattigare [poor], fn, fn:s [UN], rika [rich], Covax, Tedros, afrika [Africa], wfp [World Food Programme (WFP)], who, who:s, världshälsoorganisationen [WHO], Adhanom, Ghebreyesus, Moeti
Local COVID-19 situation at Ostrobothnia region	Salonen, Vasa (Vaasa], österbotten, österbottens [Ostrobothnia], Soite, Kinnunen, centralsjukhus, centralsjukhuset [central hospital], Kaukoranta, Nieminen, distriktet [district], soites, sjukvårdsdistriktet [health care district], österbottningarna (Ostrobothnian residents], chefsöverläkare [chief physician]
Individual sports	Världscupen (world cup], Johaug, åkare, åkarna [skier(s)], skiathlon, Ruka, Klæbo, Therese, tour, Weng, skidförbundet [ski association], Høsflot, Pärmäkoski, ski, Falun
Ice hockey leagues and teams	Edmonton, säsongen [season], nhl, nhl:s [NHL] matcher [games], målvakt [goalie], laget [team], match, matchen [match], grundserien [regural season], Montreal, rangers, slutspelet [playoffs], Columbus, isen [ice]

Source: Own elaboration.

TABLE 3 Attributes of topics in Swedish news media.

Topic	Attributes (15)—Swedish (and English equivalent)
Financial reporting	Sivler, Empir, Bultens, Gränges, ArcAroma, Doxa, Aerogel, Jagenheim, Frontwalker, Theapartment, mlinde, email., Redwood, TC Tech, Fagerhultgroup
COVID-19 virus and variations	Immuniteten, immunitet [immunity], mutationerna, mutationen, mutationer [mutations], varianten [variant], omikron, forskarna [researchers], Dillner, Mirazimi, Jalalvand, Buggert, ba., virologi [virology], Sällberg
COVID-19 abroad	Afp [AFP], nattligt [nightly], hälsodepartementet [health department], huvudstaden [capital], landet, landets [country], smittofall [cases], ap [AP], Maharashtra, Reuters, myndigheterna [authorities], hälsominister [health minister], premiärminister [prime minister], storstadsområdet [metropolitan area], Qom
Economical impact	Återhämtningen [recovery], kostnadsåtgärder [cost measures], motvindar [headwind], inflationen [inflation], tynga [weigh], dämpat [muted], tjänstesektorn [services], ekonomin [economy], utsikterna [outlook], makroekonomiska [macroeconomic], osäkerheten [uncertainty], centralbanken [central bank], inflationstryck [inflationary pressures], avmattningen [slowdown], ebita-resultatet [ebita results]
Restrictions in municipalities and elderly care	Äldreomsorgen [elderly care], kommunerna [municipalities], kommunal, kommunals [municipal], pandemilag [pandemic law], äldreboendena, äldreboenden [nursing home], liberalerna [liberals], moderaterna [moderates], hemtjänsten [home care], skyddsstopp [protection stop], centerpartiet [the Center party], arbetsmiljöverket [Work Environment Authority], karensavdraget [retirement age], skolverket [school board]
COVID-19 across Sweden	Regionen, regionerna, regionens [region], götaland [Götaland], Byfors, Skåningar, länet [country], götalandsregionen [the Götaland region], smittskyddsläkare, smittskyddsläkaren [infection control doctor], sjukvårdsdirektör [health care director], Skåne, provta [tested], vårdas [treated], platåfas [plateau phase]
Death of celebrities	Älskade [beloved], pappa [father], kärlek, kärleken [love], Wollter, sorgen [sorrow], artisten [artist], mamma [mother], dramatikern [playwright], Alsing, sonen [son], morfar [grandfather], skådespelaren [actor], sjunga [sing], saknaden [loss]
Sport leagues and teams	Matchen, matcher, matcherna [match], laget, lagets, [team], klubben, klubbens [club], truppen [squad], spelare, spelarna [player], ligan [league], spelartruppen [playing squad], anfallaren [striker], Djurgården, hemmalaget [home team]
COVID-19 and US politics	Biden, Bidens, Joe [Biden], presidenten, presidentens [president], Trump, Donald, Trumps, anhängare [supporter], demokraterna, demokraternas [Democrats], väljare [voters], vicepresident, vicepresidenten [vice president], republikanska [republican]
Stock market	Backade [backed], handlades [traded], förhandeln, förhandel [pre-trade], adderade [added], rusade [rushed], klättrade [climbed], S&P, Tesla, Exxon, retirerade [retreated], oljebolagen [oil companies], börsöppningen [stock market], nollstrecket [opening], Gamestop
Winter sport	Åkarna, åkare [skiers], världscupen, vm [World Cup], mästerskapet, mästerskap [championship], Ruka, JVM, landslaget [national team], tävlingarna [competition], truppen [squad], turneringen [tournament], OS, skidförbundet [ski association], skiathlon
Vaccine and vaccinations	Covid-19-vaccin, vaccin, vaccinet [vaccine], Keersmaecker, zeneca, zenecas astra [Zeneca], modernas [Moderna], doser, doserna [doses], svensk-brittiska [Swedish-British], Emer, läkemedelsmyndigheten [the Swedish Medicines Agency], EMA [European Medicines Agency], pfizer [Pfizer]

Source: Own elaboration.

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

national concerns. Overall, in Finland, the focus of the news coverage was predominantly on the government's management of the pandemic, virus exposure in educational institutions and the global economic situation. Wide news coverage of political decisions and business restrictions was typical in Finland throughout the pandemic, especially when new virus variants appeared and new pandemic measures were implemented. In Sweden, the coverage focused more on the national financial aspects of the pandemic, as well as on the nature and variations of the virus and the pandemic abroad.

In respect to the attributes associated with the central topics of the pandemic (second-level agenda-setting), differences in describing them are evident, showing the importance of analysing not just *what* is covered in news media, but also *how* it is covered. For instance, regarding public exposure to the virus, the news media focus in Finland was more on reporting from the perspective of exposure risks in schools and kindergartens, while in Sweden, the perspective was more on exposure risks among elderly people. Interestingly, the elderly were considered the most vulnerable group while young people were considered the least vulnerable. All news media across the two countries dedicated sufficient space to discuss the crisis origins and the role of political actors and businesses in pandemic management, evidencing on the one hand the important role of news media in crisis sense-making processes (Kozman et al., 2021) and on the other hand their propensity to ascribe crisis responsibility (Kim, 2016; Thomas et al., 2020).

5 | CONCLUSIONS AND FURTHER RESEARCH

This study explored the suitability of NLP methods to detect first- and second-level agenda-setting roles of news media in two minor language groups in a prolonged crisis. The results show that NLP methods can be utilized to analyse news media coverage of a prolonged crisis across the two Nordic countries, news outlets and time allowing detection of otherwise nuanced differences in how crisis topics are represented and framed. Using NLP in this way enables crisis managers to gain an in-depth understanding of both the media agenda (media attitudes) and the public agenda (public attitudes shaped by media attitudes) in crisis situations. Utilizing NLP methods to analyse news media coverage of the pandemic also offers broader macrolevel insights about the main crisis-related concerns of segments of publics, including who they feel are most responsible, across countries over an extended period of time. Crisis managers may subsequently make critical decisions of what should or should not be done (e.g., focus less on a topic or actor attracting negative perceptions) both during and in the immediate aftermath of a pandemic.

5.1 | Managerial implications

The practical contributions of this study are that it demonstrates the potential for NLP to assist crisis communication professionals to gain a macrounderstanding of *what* the public is thinking about (for first-level agenda-setting, see Carroll & McCombs, 2003; McCombs, 2004, 2005), *who* they feel are responsible for aspects of the crisis (for responsibility attribution see Coombs, 2021) and *how* they view critical issues based on widespread news coverage (for second-level agenda-setting see McCombs et al., 1997). As an example, NLP enables quantifiable and systematic analysis examining how key actors receiving high visibility news coverage during a crisis are portrayed (e.g., across a positive-negative attributes spectrum), including which ones are attracting higher levels of public scrutiny and which ones are considered as the highest reputational threat. These insights can provide important, real-time information to crisis communication managers about actors (including organizations) considered accountable for crisis management. NLP also offers insight into *what* media outlets have the greatest influence (for intermedia agenda-setting, see Golan, 2006) on *how* people perceive the crisis. In sum, leveraging NLP of media data for crisis communication management can reveal in real time *who* is being blamed as responsible for *what* (i.e., for aspects of a crisis as revealed by crisis topics), *how much* blame is being attributed by *whom* at *what time periods* and *where*.

Taking the news coverage of the COVID-19 pandemic in Finland and Sweden as an illustrative example, this study shows the value of NLP methods to identify, in a systematic and comparative manner, the main crisis topics embedded within news reports (i.e., first-level agenda-setting) and how these topics are described and contextualized for the public by news media professionals (i.e., second-level agenda-setting) in each country and even within sub-groups in one country (in this case between Finnish and Swedish news consumers in Finland).

Another important takeaway is related to the geolocalization of public interests, which can be helpful, for instance, to detect multicultural crisis communication concerns (Falkheimer & Heide, 2006). Second-level agenda-setting analysis can inform crisis communication managers of nuanced variations of crisis-related concerns, thus addressing the quest for more cultural differentiations in crisis communications and management (Dhanesh & Sriramesh, 2018; Siew-Yoong Low et al., 2011). As shown in this study, both Finland and Sweden news media widely discussed the impact of COVID-19 on league sports such as ice hockey, showing how important this sport is in people's daily lives. Yet, the news media covered vulnerable groups differently, with media in Sweden focusing on elderly people and media in Finland focusing on children and schoolteachers. Recognizing these patterns, authorities can calibrate their crisis efforts to support the specificities of these different groups of citizens, prioritizing specific crisis actions over others, depending on the country and/or group's main concerns. Such analyses can help to understand the evolution of crisis news reporting over time and help to manage the flow of communication to reveal what frames different actors or media outlets use when communicating about different types of crises (van der Meer, 2016). Furthermore, when done a posteriori, NLP can serve as a tool to assess the effectiveness of crisis response strategies by outlining the extent to which news media covered and treated an organization's or authority's crisis actions.

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5.2 | Theoretical implications

This study contributes to crisis framing and agenda-setting literature by exploring the suitability of NLP methods in agenda-setting studies based on minor language groups, thus supporting the refinement of Finnish and Swedish NLP research (Griciūtė et al., 2023; Hämäläinen & Alnajjar, 2021) through the inclusion of multiple and diversified news media outlets and more Nordic explorations (cf., Griciūtė et al., 2023). It thus contributes to the scholarly work on computational analysis in agenda-setting studies, showing the opportunities offered by NLP methods in analysing first- and second-level agenda-setting of large data sets, increasing the opportunities to automatize otherwise time-consuming analytical processes. This study also showcases the methodological potential to detect and analyse otherwise fragmented stakeholders and their opinions without preassigned frames, and therefore capture cultural differences that may make a difference in how individuals respond to a crisis and process instructive information (Coombs, 2021; Dhanesh & Sriramesh, 2018; Falkheimer & Heide, 2006). This, combined with the large data set, increases the possibility to compare agenda-setting across media outlets.

It also advances crisis communication research by offering empirical evidence of the diversity of crisis frames (An & Gower, 2009; Valentini & Romenti, 2011) which in early studies were constructed a priori and limitedly captured the diversity of opinions and public perceptions. The insights gained through NLP application on crisis framing point to the notion of the importance of understanding differences and nuances in how a crisis is perceived across countries and time, thus underscoring the distinct importance of analysing news media attributes and frames for planning effective crisis response messages that match the level of responsibility ascribed by the public and the overall reputational threat (Coombs, 2021) and take into consideration the specificities of media systems (Dhanesh & Sriramesh, 2018).

In this article, we have chosen to examine topic modelling, an unsupervised method which allows us to capture issue-specific frames (van der Meer, 2016), based on agenda-setting theory of news media. Future research could mix supervised and unsupervised methods and look at the coverage of selected news media during different types of crises with various levels of responsibility attribution and include other cultural perspectives as well. Large quantities of news articles should be used in the future when examining, for example, media narratives during a public health crisis and the differences between these across various news media in different countries. Triangulating the media's agenda with the public agenda, for instance in social media, and the government/state agenda (e.g., authorities' official communication) could offer a more complete picture of various actors' agendas shaping crisis perceptions of different stakeholder groups and the effectiveness of authorities' crisis response strategies in longitudinal and cross-national/group contexts. Furthermore, future crisis communication studies could examine network agenda-setting (also known as third-level agenda-setting) and could explore how salient networked crisis topics shaped the public agenda.

5.3 | Limitations

This study is not without limitations. First, the study may suffer from model and design biases. Data and pretrained NLP tools for Finnish and Swedish topic modelling are limited compared to larger language groups such as English. Like most machine learning approaches, topic modelling approaches benefit from large data sets; specifically, higher quality embeddings (and hence topics) can be extracted with more data (Mikolov et al., 2013). It is good to keep in mind that all models make assumptions; in the case of Top2vec, topics are considered to be clusters of similar documents. We additionally assumed that topics contain at least 20 documents, which could have suppressed some infrequent topics. The topic models used for studying news media in Finland and Sweden were constructed by one of the researchers based on the research questions, while the labelling of the generated topics was performed by another researcher and evaluated by the other authors. To reduce the subjectivity in labelling, as this is an interpretive process based on the words and/or the documents (van der Meer, 2016), all researchers discussed the labels referring to second-level agenda setting. This increased the validity of the process. Second, this study did not examine agenda-setting effects between media types (i.e., news outlets and individual journalists). This would be interesting to explore in future studies, for example, to observe how coverage from some types of media (e.g., elite media) shape how other types of media represent and treat certain crisis topics. Third, although the data of this study includes a large number of news reports published by Finnish, Finnish-Swedish and Swedish news media, the data was limited to online news articles about COVID-19 published in Finland and Sweden during the period February 2020 and June 2022 and thus does not contain articles published behind digital paywalls or contain multimedia content such as images and videos. Also, the materials of each news media outlet have their own differences. For instance, very few hits from *Dagens Nyheter* compared to the size of the news outlet. As we trained the topic models on the news articles gathered in this study, the models might be affected by selection bias due to the limitations in the data gathering process (Hovy & Prabhumoye, 2021). NER results can contain multiple different strings referring to an entity (e.g., 'WHO', 'WHO:n', and 'World Health Organization (WHO)'). We attempted to merge these entries automatically, but this was not always possible. There are also differences between countries. For Finland, the data covers only the three largest news outlets and their articles which we had free access to. As Finland is a bilingual country, we concentrated only on three Finnish-language news outlets and included two Finnish-Swedish-language news outlets to compare crisis topics. Including more Finnish-language news outlets would have given more broader insight into Finnish topics. For technical reasons, we limited the attributes analysis to whole language-related data. Analysing attributes by news media or year would have given more detailed insights about second-level agenda-setting of news media in both countries.

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ACKNOWLEDGEMENTS

This study is a part of the project 'COVID-19 Voices in Finnish News Media in the Global Context' founded by the Helsingin Sanomat Foundation, Grant Number: 202000087.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Due to copyright reasons, the data set of this study can be access only upon request. The natural language processing codes are publicly available in the project university repository space.

ORCID

Markus Mykkänen http://orcid.org/0000-0002-7044-9263
Anri Patron http://orcid.org/0000-0002-4728-0391
Mark Badham http://orcid.org/0000-0003-4521-5294
Chiara Valentini http://orcid.org/0000-0003-0752-9639

REFERENCES

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, *3*, 42. https://doi.org/10.3389/frai.2020.00042

Ali, S. M. A., & Sherman-Morris, K. (2023). Pandemic and health reporting: A content analysis of New York Times coverage of COVID-19 from January 01, 2020, to August 31, 2022. *Social Sciences & Humanities Open*, *8*(1), 100739. https://doi.org/10.1016/j.ssaho.2023.100739

An, S. K., & Gower, K. K. (2009). How do the news media frame crises? A content analysis of crisis news coverage. *Public Relations Review*, *35*(2), 107–112. https://doi.org/10.1016/j.pubrev.2009.01.010

Andreasson, U. (2017). *Trust—The Nordic gold*. Nordic Council of Ministers.

Angelov, D. (2020). Top2vec: Distributed representations of topics. *arXiv*, 1–25. https://doi.org/10.48550/arXiv.2008.09470

Badham, M. (2019). Four news media roles shaping agenda-building processes. In F. Frandsen, W. Johansen, R. Tench, & S. Romenti (Eds.), *Big ideas in public relations research and practice*. Advances in Public Relations and Communication Management (Vol. 4, pp. 163–176). Emerald. https://doi.org/10.1108/S2398-39142019000000401

Blair, S. J., Bi, Y., & Mulvenna, M. D. (2020). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, *50*(1), 138–156. https://doi.org/10.1007/s10489-019-01438-z

Boin, A., 't Hart, P., & McConnell, A. (2009). Crisis exploitation: Political and policy impacts of framing contests. *Journal of European Public Policy*, *16*, 81–106.

Boumans, J. W., & Trilling, D. (2018). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars, In M. Karlsson & H. Sjovaag (Eds.) *Rethinking research methods in an age of digital journalism* (pp. 8–23). Routledge.

Bowen, S. A., & Zheng, Y. (2015). Auto recall crisis, framing, and ethical response: Toyota's missteps. *Public Relations Review*, *41*, 40–49. https://doi.org/10.1016/j.pubrev.2014.10.017

De Bruyne, L., van der Meer, T. G. L. A., De Clercq, O., & Hoste, V. (2024). Using state-of-the-art emotion detection models in a crisis communication context. *Computational Communication Research*, *6*(1), 1–45.

Burggraaff, C., & Trilling, D. (2020). Through a different gate: An automated content analysis of how online news and print news differ. *Journalism*, *21*(1), 112–129. https://doi.org/10.1177/1464884917716699

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu, *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science* (Vol. 7819, pp. 160–172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

Carroll, C. E., & McCombs, M. (2003). Agenda-setting effects of business news on the public's images and opinions about major corporations. *Corporate Reputation Review*, *6*, 36–46. https://doi.org/10.1057/palgrave.crr.1540188

Chen, Y., Peng, Z., Kim, S.-H., & Choi, C. W. (2023). what we can do and cannot do with topic modeling: A systematic review. *Communication Methods and Measures*, *17*(2), 111–130. https://doi.org/10.1080/19312458.2023.2167965

Cobb, R. W., & Elder, C. D. (1983). *Participation in American politics: The dynamics of agenda building*. Johns Hopkins Press.

Colleoni, E., Romenti, S., Valentini, C., Badham, M., Choi, S. I., Kim, S., & Jin, Y. (2022). Does culture matter?: Measuring cross-country perceptions of CSR communication campaigns about COVID-19. *Sustainability*, *14*(2):889. https://doi.org/10.3390/su14020889

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. *Corporate Reputation Review*, *10*, 163–176. https://doi.org/10.1057/palgrave.crr.1550049

Coombs, W. T. (2012). Parameters for crisis communication. In W. T. Coombs & S. J. Holladay (Eds.), *The handbook of crisis communication* (pp. 17–53). Wiley-Blackwell.

Coombs, T. W. (2021). *Ongoing crisis communication: Planning, managing and responding* (6th ed.). Sage.

Deng, Q., Gao, Y., Wang, C., & Zhang, H. (2020). Detecting information requirements for crisis communication from social media data: An interactive topic modeling approach. *International Journal of Disaster Risk Reduction*, *50*, 101692. https://doi.org/10.1016/j.ijdrr.2020.101692

Dhanesh, G. S., & Sriramesh, K. (2018). Culture and crisis communication: Nestle India's Maggi noodles case. *Journal of International Management*, *24*(3), 204–214. https://doi.org/10.1016/j.intman.2017.12.004

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

Falkheimer, J., & Heide, M. (2006). Multicultural crisis communication: Towards a social constructionist perspective. *Journal of Contingencies and Crisis Management*, *14*(4), 180–189. https://doi.org/10.1111/j.1468-5973.2006.00494.x

Gallagher, R. J., Reing, K., Kale, D., & Ver Steeg, G. (2017). Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Transactions of the Association for Computational Linguistics*, *5*, 529–542. https://doi.org/10.1162/tacl_a_00078

Ghanem, S. (1997). Filling in the tapestry: The second level of agenda setting, In M. D. McCombs, D. L. Shaw, & D. Weaver (Eds.) *Communication and democracy: Exploring the intellectual frontiers in agenda-setting theory* (pp. 3–14). Lawrence Erlbaum Associates.

Gigliotti, R. A. (2020). The perception of crisis, the existence of crisis: Navigating the social construction of crisis. *Journal of Applied Communication Research*, *48*(5), 558–576. https://doi.org/10.1080/00909882.2020.1820553

Golan, G. (2006). Inter-media agenda setting and global news coverage. *Journalism Studies*, *7*(2), 323–333. https://doi.org/10.1080/14616700500533643

Griciūtė, B., Han, L., & Nenadic, G. (2023). Topic modelling of Swedish newspaper articles about coronavirus: A case study using Latent Dirichlet Allocation method, *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)* (pp. 627–636). IEEE.

Grootendorst, M. (2021). Interactive topic modeling with BERTopic. Towards Data Science. https://towardsdatascience.com/interactive-topic-modeling-with-bertopic-1ea55e7d73d8

Guo, L., & McCombs, M. (2011). *Network agenda setting: A third level of media effects*. Paper presented at the ICA.

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge University Press.

Hämäläinen, M., & Alnajjar, K. (2021). The current state of Finnish NLP. *arXiv*, 1–8. https://doi.org/10.48550/arXiv.2109.11326

Happer, C., & Philo, G. (2013). The role of the media in the construction of public belief and social change. *Journal of Social and Political Psychology*, *1*(1), 321–336. https://doi.org/10.5964/jspp.v1i1.96

Heidenreich, T., Lind, F., Eberl, J. M., & Boomgaarden, H. G. (2019). Media framing dynamics of the 'European refugee crisis': A comparative topic modelling approach. *Journal of Refugee Studies*, *32*(1), i172–i182. https://doi.org/10.1093/jrs/fez025

Hovden, J. F., & Väliverronen, J. (2021). Nordic journalists' conceptual roles and perceived influences: A European and inter-Nordic comparison. *Nordicom Review*, *42*(1), 141–161. https://doi.org/10.2478/nor-2021-0034

Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and linguistics compass*, *15*(8), e12432.

Iannarino, N. T., Veil, S. R., & Cotton, A. J. (2015). Bringing home the crisis: How US Evening News Framed the 2011 Japan Nuclear Crisis. *Journal of Contingencies and Crisis Management*, *23*, 169–181. https://doi.org/10.1111/1468-5973.12068

Jacobs, T., & Tschötschel, R. (2019). Topic models meet discourse analysis: A quantitative tool for a qualitative approach. *International Journal of Social Research Methodology*, *22*(5), 469–485. https://doi.org/10.1080/13645579.2019.1576317

Jockers, M. L., & Thalken, R. (2020). Topic modeling. In M. L. Jockers & R. Thalken (Eds.), *Quantitative methods in the humanities and social sciences. Text analysis with R* (pp. 211–235). Springer International Publishing. https://doi.org/10.1007/978-3-030-39643-5_17

Johansson, B., & Odén, T. (2018). Struggling for the Upper Hand: News sources and crisis communication in a digital media environment. *Journalism Studies*, *19*(10), 1489–1506. https://doi.org/10.1080/1461670X.2017.1279980

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2nd ed.). Prentice Hall.

Kanerva, J., Ginter, F., Miekka, N., Leino, A., & Salakoski, T. (2018). Turku neural parser pipeline: An end-to-end system for the CoNLL 2018 shared task, In D. Zeman & Jan Hajič (Eds.) *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing From Raw Text to Universal Dependencies, Brussels, Belgium, October 2018* (pp. 133–142). Association for Computational Linguistics. https://doi.org/10.18653/v1/K18-2013

Kanerva, J., Ginter, F., & Salakoski, T. (2021). Universal lemmatizer: A sequence-to-sequence model for lemmatizing universal dependencies treebanks. *Natural Language Engineering*, *27*(5), 545–574. https://doi.org/10.1017/S1351324920000224

Kim, Y. (2016). Understanding publics' perception and behaviors in crisis communication: Effects of crisis news framing and publics' acquisition, selection, and transmission of information in crisis situations. *Journal of Public Relations Research*, *28*(1), 35–50. https://doi.org/10.1080/1062726X.2015.1131697

Kim, S., Choi, S. I., Valentini, C., Badham, M., & Jin, Y. (2023). How motivation to reduce uncertainty predicts COVID-19 behavioral responses: Strategic health communication insights for managing an ongoing pandemic. *American Behavioral Scientist*. https://doi.org/10.1177/00027642231164051

Kim, S.-H., Han, M., Choi, D.-H., & Kim, J.-N. (2012). Attribute agenda setting, priming and the media's influence on how to think about a controversial issue. *International Communication Gazette*, *74*(1), 43–59. https://doi.org/10.1177/1748048511426991

Kim, Y., Kim, S., Jaimes, A., & Oh, A. (2014). A computational analysis of agenda setting, In C. Chung (Ed.) *Proceedings of the 23rd International Conference on World Wide Web* (pp. 323–324). ACM.

Kim, Y., Kim, Y., & Zhou, S. (2017). Theoretical and methodological trends of agenda-setting theory: A thematic analysis of the last four decades of research. *The Agenda Setting Journal*, *1*(1), 5–22.

Korenčić, D., Ristov, S., & Šnajder, J. (2015). Getting the agenda right: measuring media agenda using topic models, *Proceedings of the 2015 Workshop on Topic Models: Post-Processing and Applications* (pp. 61–66). ACM.

Kozman, C., Tabbara, R., & Melki, J. (2021). The role of media and communication in reducing uncertainty during the Syria war. *Media and Communication*, *9*(4), 297–308. https://doi.org/10.17645/mac.v9i4.4352

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents, In E. Xing & T. Jebara (Eds.) *International Conference on Machine Learning* (pp. 1188–1196). PMLR. https://doi.org/10.48550/arXiv.1405.4053

Leydesdorff, L., & Hellsten, I. (2006). Measuring the meaning of words in contexts: An automated analysis of controversies about 'Monarch butterflies', 'Frankenfoods' and 'stem cells'. *Scientometrics*, *67*, 231–258.

Lim, J., & Seo, H. (2009). Frame flow between government and the news media and its effects on the public: Framing of North Korea. *International Journal of Public Opinion Research*, *21*(3), 204–223.

Lindstedt, N. C. (2019). Structural topic modeling for social scientists: A brief case study with social movement studies literature, 2005–2017. *Social Currents*, *6*(4), 307–318. https://doi.org/10.1177/2329496519846505

Lu, G., Businger, M., Dollfus, C., Wozniak, T., Fleck, M., Heroth, T., Lock, I., & Lipenkova, J. (2023). agenda-setting for COVID-19: A study of large-scale economic news coverage using natural language processing. *International Journal of Data Science and Analytics*, *15*(3), 291–312. https://doi.org/10.1007/s41060-022-00364-7

MacNamara, J. R. (2005). Media content analysis: Its uses, benefits and best practice methodology. *Asia Pacific Public Relations Journal*, *6*(1), 1–34. https://doi.org/10.3316/ielapa.200705762

Mahon, J. F., Heugens, P. P. M. A. R., & McGowan, R. A. (2018). Blending issues and stakeholders: In pursuit of the elusive synergy. *Journal of Public Affairs*, *18*, e1635. https://doi.org/10.1002/pa.1635

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

McCombs, M. E. (2004). *Setting the agenda: The mass media and public opinion*. Polity Press.

McCombs, M. (2005). A look at agenda-setting: Past, present and future. *Journalism Studies*, *6*(4), 543–557. https://doi.org/10.1080/14616700500250438

McCombs, M., Llamas, J. P., Lopez-Escobar, E., & Rey, F. (1997). Candidate images in Spanish elections: Second-level agenda-setting effects. *Journalism & Mass Communication Quarterly*, *74*(4), 703–717. https://doi.org/10.1177/107769909707400404

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, *36*(2), 176–187.

van der Meer, T. G. L. A. (2016). Automated content analysis and crisis communication research. *Public Relations Review*, *42*(5), 952–961. https://doi.org/10.1016/j.pubrev.2016.09.001

van der Meer, T. G. L. A., Verhoeven, P., Beentjes, J. W. J., & Vliegenthart, R. (2017). Disrupting gatekeeping practices: Journalists' source selection in times of crisis. *Journalism*, *18*(9), 1107–1124. https://doi.org/10.1177/1464884916648095

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*, 1–12. https://doi.org/10.48550/arXiv.1301.3781

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality, In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.) *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems* (pp. 3111–3119). ACM.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimising semantic coherence in topic models. In R. Barzilay, & M. Johnson (Eds.), *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262–272). Association for Computational Linguistics.

Murphy, J. (1983). Words: What goes with what?, *Paper presented at the 17th Annual Convention of Teachers of English to Speakers of Other Languages, Toronto, Ontario, March 15–20*.

Van Nguyen, M., Lai, V. D., Veyseh, A. P. B., & Nguyen, T. H. (2021). Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. *arXiv*, 80–90. https://doi.org/10.48550/arXiv.2101.03289

Öhman, S., Giritli Nygren, K., & Olofsson, A. (2016). The (un)intended consequences of crisis communication in news media: A critical analysis. *Critical Discourse Studies*, *13*(5), 515–530. https://doi.org/10.1080/17405904.2016.1174138

Ophir, Y. (2018). Coverage of epidemics in American newspapers through the lens of the crisis and emergency risk communication framework. *Health Security*, *16*(3), 147–157. https://doi.org/10.1089/hs.2017.0106

Page, T. G., Zhou, A., & Capizzo, L. W. (2023). Finding the path beyond reputation repair: A structural topic modeling analysis of the crisis communication paradigm in public relations. *Public Relations Review*, *49*(4), 102349. https://doi.org/10.1016/j.pubrev.2023.102349

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information*, *2*(1–2), 1–135.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. In A. Celikyilmaz & T.-H. Wen (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 101–108). Association for Computational Linguistics.

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora, *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). ELRA.

Reunanen, E., Alanne, N., Huovinen, T., Järvi, U., Nevalainen, R., Puolimatka, R., & Vehkasalo, V. (2023). *Uutismedia verkossa 2023. Reuters-Institute Digital News Report-Suomen maaraportti* [News media online 2023. Reuters Institute's Digital News Report-Finland country report].

Rogers, J., Müller, A., Daulton, F. E., Dickinson, P., Florescu, C., Reid, G., & Stoeckel, T. (2021). The creation and application of a large-scale corpus-based academic multi-word unit list. *English for Specific Purposes*, *62*, 142–157. https://doi.org/10.1016/j.esp.2021.01.001

Seeger, M. W., Sellnow, T. L., & Ulmer, R. R. (2003). *Communication and organizational crisis*. Praeger.

Siew-Yoong Low, Y., Varughese, J., & Pang, A. (2011). Communicating crisis: how culture influences image repair in Western and Asian governments. *Corporate Communications: An International Journal*, *16*(3), 218–242. https://doi.org/10.1108/13563281111156880

Statistics Finland. (2023). Population structure on 31 December. https://www2.stat.fi/tup/suoluk/suoluk_vaesto_en.html#Population%20structure%20on%2031%20December

SV, P., & Ittamalla, R. (2021). An analysis of attitude of general public toward COVID-19 crises—Sentimental analysis and a topic modeling study. *Information Discovery and Delivery*, *49*(3), 240–249. https://doi.org/10.1108/IDD-08-2020-0097

Thomas, T., Wilson, A., Tonkin, E., Miller, E. R., & Ward, P. R. (2020). How the media places responsibility for the COVID-19 pandemic—An Australian media analysis. *Frontiers in Public Health*, *8*, 483. https://doi.org/10.3389/fpubh.2020.00483

Toivanen, P., Huhtamäki, J., Valaskivi, K., & Tikka, M. (2020). Aihemallinnus hybridin mediatapahtuman ja merkitysten kierron tutkimuksessa [Topic modeling in studying hybrid media events and circulation of meanings]. *Media & viestintä*, *43*(1), 1–20. https://doi.org/10.23983/mv.91078

Tomasi, S., Kumble, S., Diddi, P., & Parolia, N. (2023). The framing of initial COVID-19 communication: Using unsupervised machine learning on press releases. *Business and Society Review*, *128*(3), 515–531. https://doi.org/10.1111/basr.12323

Turk, J. V., Jin, Y., Stewart, S., Kim, J., & Hipple, J. R. (2012). Examining the interplay of an organization's prior reputation, CEO's visibility, and immediate response to a crisis. *Public Relations Review*, *38*, 574–583. https://doi.org/10.1016/j.pubrev.2012.06.012

Valentini, C., & Badham, M. (2023). Public trust in governments' communicating with intermediaries: Finnish and Swedish Governments during the COVID-19 pandemic. In R. Tench, J. Meng, & Á. Moreno (Eds.), *Strategic communication in a global crisis: National and international responses to the COVID-19 Pandemic* (pp. 165–183). Routledge. https://doi.org/10.4324/9781003184669-16

Valentini, C., & Romenti, S. (2011). The press and Alitalia's 2008 crisis: Issues, tones, and frames. *Public Relations Review*, *37*(4), 360–365. https://doi.org/10.1016/j.pubrev.2011.07.002

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. https://doi.org/10.1016/j.is.2020.101582

Waddock, S. A., & Mahon, J. F. (1991). Corporate social performance revisited: Dimensions of efficacy, effectiveness, and efficiency. In J. E. Post (Ed.), *Research in corporate social performance and policy* (Vol. 12, pp. 231–262). JAI Press.

Wang, M., & Mengoni, P. (2020). How pandemic spread in news: Text analysis using topic model, In J. He, W. van den Heuvel, & J. Yearwood (Eds.) *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 764–770). IEEE. https://doi.org/10.1109/WIIAT50758.2020.00118

Wanta, W., & Alkazemi, M. F. (2017). Agenda-setting: History and research tradition. In P. Rössler, C. A. Hoffner, & L. Van Zoonen, *The International Encyclopedia of Media Effects* (pp. 1–14). John Wiley & Sons. https://doi.org/10.1002/9781118783764.wbieme0030

Weaver, D. H. (2007). Thoughts on agenda setting, framing, and priming. *Journal of Communication*, *57*(1), 142–147. https://doi.org/10.1111/j.1460-2466.2006.00333.x

Wei, X., Xin, L., & Yinhong, G. (2003). Document clustering based on non-negative matrix factorization. In J. Callan (Ed.), *Special Issue of the SIGIR Forum, SIGIR 2003: Proceedings of the Twenty-Sixth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada, July 28 to August 1, 2003* (pp. 267–273). ACM Press.

Westlund, O. (2023). News media online 2023. Reuters Institute's Digital News Report—Sweden country report. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/sweden

Wigley, S., & Fontenot, M. (2011). The Giffords shootings in Tucson: Exploring citizen-generated versus news media content in crisis management. *Public Relations Review*, *37*(4), 337–344. https://doi.org/10.1016/j.pubrev.2011.07.004

Wray, R. J., Becker, S. M., Henderson, N., Glik, D., Jupka, K., Middleton, S., Henderson, C., Drury, A., & Mitchell, E. W. (2008). Communicating with the public about emerging health threats: Lessons from the pre-event message development project. *American Journal of Public Health*, *98*(12), 2214–2222. https://doi.org/10.2105/ajph.2006.107102

Wright, C. R. (1960). Functional analysis and mass communication. *Public Opinion Quarterly*, *24*, 605–620.

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wu, K. (2021). Agenda-setting in cross-national coverage of COVID-19: An analysis of elite newspapers in US and China with topic modeling. *Online Journal of Communication and Media Technologies*, *11*(4), e202116. https://doi.org/10.30935/ojcmt/11083

Xu, Z. (2023). Examining US newspapers' partisan bias in COVID-19 news using computational methods. *Communication Studies*, *74*(1), 78–96. https://doi.org/10.1080/10510974.2023.2169729

Zhang, W., Hu, L., & Park, J. (2023). Politics go "viral": A computational text analysis of the public attribution and attitude regarding the COVID-19 crisis and governmental responses on Twitter. *Social Science Computer Review*, *41*(3), 790–811. https://doi.org/10.1177/08944393211053743

How to cite this article: Mykkänen, M., Patron, A., Badham, M., & Valentini, C. (2024). Leveraging NLP for crisis communication management: A case study of news media analysis of the COVID-19 pandemic in two Nordic countries. *Journal of Contingencies and Crisis Management*, *32*, e12611. https://doi.org/10.1111/1468-5973.12611

14685973, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-5973.12611 by Leeds Beckett University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License