

Special Feature

Lasting Themes Beyond Vocabulary Shifts: Human Rights Discourse in South Korea, 1990–2010

Eunkyung SONG



The Review of Korean Studies Volume 29 Number 1 (June 2026): 59–95

doi: 10.25024/review.2026.29.1.003

©2026 by the Academy of Korean Studies.

Introduction

Recent scholarship on human rights discourse has added a new methodological layer by adopting computational text analysis tools (Bagozzi and Berliner 2018; Fariss et al. 2015; J. Kim 2021; Koo and Choi 2019; Murdie, Davis, and Park 2020; Park, Murdie, and Davis 2019). These approaches have refined existing datasets designed to assess the impact of human rights discourse on improvements in human rights conditions (Fariss et al. 2015; Murdie, Davis, and Park 2020). They also have inspired new inquiries into the substantive content of human rights discourse, both broadly and with respect to vulnerable groups, such as sexual minorities, refugees, and migrant workers (C. Kim et al. 2022; J. Kim 2021; Oh and Shin 2024; Park, Murdie, and Davis 2019). As in earlier research, texts produced and circulated by non-state actors—particularly human rights organizations (HROs) and media outlets at both the global and domestic levels—have received the greatest scholarly attention (Fariss et al. 2015; Hendrix and Wong 2013; Koliev and Lebovic 2018; Koo and Choi 2019; Meernik et al. 2012; Murdie, Davis, and Park 2020; Murdie and Davis 2012; Murdie and Peksen 2014; Ovsiovitch 1993; Park, Murdie, and Davis 2019; Ramos, Ron, and Thomas 2007).

Building on these advances, this study seeks to make both theoretical and methodological contributions by examining the dynamics of human rights discourse in South Korea (hereafter Korea) between 1990 and 2010. Combining computational text analysis with insights from social movement scholarship that intersects with cultural and social network analyses (Benford and Snow 2000; Fuhse and Mützel 2011; McLean 2017; Mische 2003; Steinberg 1998, 1999), this study conceptualizes human rights discourse as a joint product of communication among diverse stakeholders, whose distinctive objectives and institutional positions shape not only the contents of their claims but also how they navigate various issues. By incorporating the insight that discourse has at least two properties—content and performer—this study asks (1) how Korean human rights discourse is represented by topics—i.e., what is discussed—whose substance is updated at varying speeds and whose manifestation differs in continuity; (2) how stakeholders tend to connect those topics differently across time; and (3) whether such patterns, if any, become salient when stakeholders aim to promote issues that are semantically similar.

This relational approach is theorized in two ways. First, it traces multiple

stakeholders of human rights discourse, specifically by including the state as an active contributor to discursive dynamics. Rather than positioning the state as analytically external to human rights discourse advanced by non-state actors—as is common in prior research (Keck and Sikkink 1998; Murdie, Davis, and Park 2020; Tsutsui and Smith 2019)—this study recasts government human rights institutions as participants that generate institutional interpretations and recommendations in response to claims of abuse and discrimination. Secondly, this study proposes an analytic approach that combines how different human rights themes are (dis)connected and how different stakeholders enact them in distinct ways. In so doing, the study represents human rights discourse as a product of (dis)connected topics that is “Realized” in association with semantic similarity, “Unrealized” despite semantic similarity, and “Surprising” in their semantic dissimilarity.

Empirically, this study compiles a novel corpus of about 39,000 documents produced by human rights and civic organizations, news outlets, and the National Human Rights Commission of the Republic of Korea (NHRCK) from 1990 to 2010. BERTopic models (Grootendorst 2022) are used to classify the documents into topics by semantic similarity, which then fed to downstream analysis using multiple measures, such as topic drift (how much a topic’s vocabularies change), topic span (how long a topic survives), Pointwise Mutual Information (PMI, how stakeholders bridge topics) (Church and Hanks 1990), and social network analysis, which is used to measure how semantically (dis) similar topics end up in being connected or not.

Altogether, this paper argues that Korean human rights discourse presents great persistence in maintaining issues, whose vocabularies are either stable or renewed continuously with the following main findings. First, despite noticeable differences among the stakeholders—human rights organizations (HROs), news outlets, and the NHRCK—they all engage with topics persistently regardless of their contents and use more settled or unstable vocabularies. The NHRCK slightly deviates from this tendency, as it has put more emphasis on transient topics to capture the concrete nature of human rights abuse cases it adjudicates. Second, topics that differ in drift-span index tend not to be connected within the same document. Third, the stakeholders do not always connect semantically similar topics in a document. Some are realized, while some are not. Also, dissimilar topics are connected—a phenomenon that tends to occur with persistent topics that lack a settled set of vocabularies.

Human Rights Texts and Its Producers

Scholarship on human rights discourse has largely focused on what is being written and mentioned as well as who is performing activities regarding human rights. Various international human rights texts, mainly from the United Nations and its member states, operate at the global level, though there have rarely been unanimous interpretations of these texts (Jung 2024). In addition, numerous active stakeholders are involved in human rights discourse. International human rights organizations (HROs) have been prominent players in the production and circulation of information—particularly information that can be easily screened by authoritative regimes—by publishing annual monitoring reports, urgent action calls, and press releases (Fariss 2014; Hendrix and Wong 2013; Meernik et al. 2012; Murdie, Davis, and Park 2020; Park, Murdie, and Davis 2019). Research has also examined reports authored by international non-governmental organizations such as the International Labour Organization (ILO) or United Nations agencies (Koliev and Lebovic 2018; Lebovic and Voeten 2006, 2009) as well as annual country reports compiled by the US State Department (Bagozzi and Berliner 2018). News reports covered by both domestic and international outlets have been regarded as key channels through which human rights discourse can gain enough leverage to contribute substantively to the improvement of human rights (Kang and Lee 2019; C. Kim et al. 2022; J. Kim 2021; S. Kim 2019; Koo and Choi 2019; Murdie and Davis 2012; Oh and Shin 2024).

Recently, researchers have proposed more sophisticated automation of coding procedures to turn human rights texts into data by adopting computational text analysis methods (Fariss et al. 2015; Murdie, Davis, and Park 2020). To enhance and scale consistent and reliable textual data construction, researchers have introduced novel approaches that treat the words in the text as data points. Fariss et al. (2015) compile multisource human rights texts authored by multiple bodies and represent them as a document-term matrix, with documents in the rows and all terms to be analyzed in the columns, which is essential for downstream quantitative text analysis. Murdie, Davis, and Park (2020) probe more specifically how the operationalization of naming and shaming could be improved. Instead of reducing human rights texts to events data that focus only on who shames which country and for what, they propose ways to measure the details of such mechanisms by, for instance, considering the

intensity of the shaming.

Beyond data management, researchers have also posed new inquiries that zero in on the emergence and transformation of human rights themes at the international level. Park, Murdie, and Davis (2019) analyze two decades' worth of HRO documents to trace convergence and divergence in human rights topics over time. They demonstrate, among other things, how international justice and LGBTQ rights emerge in tandem over a specific time period. Bagozzi and Berliner (2018) apply structural topic models to US State Department Country Reports on Human Rights Practices from 1977 to 2012, identifying temporal and geographic variations in human rights themes. After specifying 15 topics—e.g., killings/disappearances, labor rights, terrorism/torture, and economic systems—they examine how these topics have been discussed differently.

Finally, another thread of emerging scholarship has started to examine a discursive landscape of the media in local contexts. Koo and Choi (2019) investigate Korean newspaper coverage from 1990 to 2020, focusing on partisan biases in the coverage of UN-defined human rights issues. Whereas Koo and Choi aim to delineate an overall typology of human rights discourse over time, other studies have explored how specific social groups or issues have been portrayed in the media (Kang and Lee 2019) as well as broader issue domains (J. Kim 2021; Oh and Shin 2024).

Missing Pieces

The increasing attention to human rights texts has advanced our understanding of how they have impacted human rights conditions, but a significant portion of the literature focuses on single-source data, with the most widely studied data sources being international HROs and global news media. Human rights discourse is widely studied at the international level because severe human rights abuses are extremely difficult to address and correct via direct communication between the abuser and the abused, largely due to the fact that the former—often the government of a country—can become more oppressive and violent in reaction to monitoring or criticisms. For this perspective, documents circulated by international HROs and global news outlets serve as primary data sources.

This approach, however, has underrepresented how different local stakeholders engage with human rights discourse as they promote and respond to political and social changes in their region. In Korea, the 1990s marked a

critical turning point for civil society, as the primary focus of social movements expanded from democratization to include environmental protection, women's rights, and institutional reform (Lee 2011; Moon 2005, 2024; Shin and Chang 2011). Unlike earlier democratization movements that prioritized civil and political rights above all else, post-1990s activism advanced protections for previously marginalized groups' rights, often within a human rights framework (Kwon Kim and Cho 2011; Seo 2005; Seokyo Humanities and Social Sciences Research Lab 2017). It became crucial for emerging organizations that had distinct agendas to establish new rhetoric and perspectives while still remaining connected to existing networks (S. Kim 2019; Moon 2005, 2024). These shifts were also reflected in the media, which increasingly covered civil society organizations and their activities (Kern and Nam 2011). In addition, the establishment of the NHRCK in 2001 was likewise inseparable from this broader political climate, which experienced more general awareness of human rights, newer rhetoric, and a novel concept of institutional agency. In other words, human rights discourse at the domestic level deserves more scholarly attention.

This perspective also challenges a key facet of the existing literature, which tends to reify the state as a singular perpetrator of human rights violations. This tendency reflects two claims with a similar attitude toward the state in terms of the politics of human rights. On the one hand, it positions the state as the culprit by default, as transnational non-state actors are promoted as the primary drivers of human rights protections against abusive governments (Keck and Sikkink 1998). On the other hand, it is rooted in critiques of human rights as international frameworks that lack state-level enforcement mechanisms (Shafir 2004). As a result, the state appears in the literature either as an active perpetrator or a passive political subject with limited authority. Yet, the state is not a monolithic actor; it comprises a constellation of institutions that may operate inconsistently (Evans, Rueschemeyer, and Skocpol 1985; Morgan and Orloff 2017). From this perspective, while some state institutions may violate rights, others may actively participate in articulating, institutionalizing, or advancing human rights values. As Koo (2011) demonstrates, debates over institutional independence and authority between the state and HROs culminated in the establishment of the NHRCK in Korea, illustrating how the state can function simultaneously as both a subject and a producer of human rights discourse.

Research in this area lacks sufficient discussion of the dynamics of human

rights discourse. Innovative computational approaches have enabled researchers to distill a large corpus of documents into a concise set of topics represented by keyword clusters and trace their longitudinal trends (Bagozzi and Berliner 2018; J. Kim 2021; Park, Murdie, and Davis 2019). However, these trends largely have not considered either the relationships between the key actors or how the identified topics are interconnected vis-à-vis the underlying influences that actors exert on one another (Fuhse and Mützel 2011; Karell and Freedman 2020). Understanding discourse formation as a fundamental part of the interactions between actors engaging in the discourse allows us to ask new questions, such as which human rights issues have persisted longer or received disproportionate attention, or how such persistence can be explained by the ways in which key actors engaged with one another.

The Current Study

Drawing on social movement scholarship, this study conceptualizes human rights discourse as a semantic space jointly constituted by multiple actors who engage with a range of issues, problems, ideas, and claims (Steinberg 1998, 1999). Steinberg (1998, 1999) underscores that communicative practices by social movement organizations rarely occur in a vacuum. Instead, such organizations attentively aim to engage with ongoing or mainstream narratives when promoting their claims. In such dialogic processes in which claimants, collaborators, supporters, and opponents respond to and reinterpret one another's arguments, meanings of the issue itself and claims are formed. In a similar vein, Mische (2003) argues that the meaning of communicative practices is embedded in their interactional contexts. With more emphasis on in-person interactions, Mische points out that various moments of interactions between activists—whether they seem directly related to the claims being made or not—often contribute to the meaning-making behind what is being communicated.

Applied to the current study, we can examine how their objectives and institutional positions shape their communication styles. Local HROs, for example, respond promptly to emerging social problems, framing them as rights-based issues and adopting new perspectives to reframe their existing concerns (Chua 2015; Șerban Rosen and Yoon 2009; Tsutsui 2017). It means that local HROs can quickly integrate new language to describe emerging

abuses within their own existing framework of concrete claims. However, in doing so, they also establish links to foundational human rights concerns rooted in historical struggles, which are surrounded by familiar and settled rhetoric. In Korea, human rights activism is deeply anchored in resistance to military authoritarianism, which systematically violated the freedom of conscience, assembly, bodily integrity, and legal protection in modern Korean history (Chang 2015; Moon 2005). In addition, because the 1970s was marked by rapid economic development alongside severe labor union oppression, labor rights claims played a pivotal role in connecting domestic organizations to international HROs (Hwang 2022). These longstanding issues and perspectives formed the strong basis for a new branch in labor rights movements that underlined immigrant worker issues since the early 1990s.

Media outlets are guided by distinct institutional logics. News organizations routinely report on long-standing human rights concerns as well as emergent issues that could capture public attention. In this way, these outlets can fundamentally shape HROs' claims and transmit their appeals to broader audiences (Apodaca 2007; Heinze and Freedman 2010; Pruce and Budabin 2016; Winston and Pollock 2016). Media coverage can also function as a channel for critique directed at state institutions, political authorities, and civil society actors. However, media attention is often episodic and shaped by perceived newsworthiness, geopolitical relevance, and national interest rather than sustained advocacy (Gordon and Berkovitch 2007; Ovsiovitch 1993). Though earlier studies have treated news outlets as platforms through which HROs promote their agendas, news organizations are neither neutral nor transparent spaces (Koo and Choi 2019; Pruce and Budabin 2016). Indeed, their respective political biases may lead to human rights issues being highlighted differently, even if the outlets pursue purported newsworthy stories. As a result, even when the media addresses topics endorsed by HROs, engagement is inconsistent over time.

Domestic human rights institutions contribute to discourse in yet another manner. Institutions such as the NHRCK occupy a distinctive position due to their quasi-legal authority. As Koo (2014) shows, debates surrounding the establishment of the NHRCK reflected competing views on institutional independence, state responsibility, and the scope of legitimate human rights claims. Unlike HROs and media outlets, this institution acts primarily within the confines of its legally defined authority, such as investigating individual

complaints and issuing policy recommendations. Also, the position of domestic human rights institutions tends to be expressed within institutional and legal language by referring to related laws or policies. In this light, the discursive contributions can be highly narrow and specific in scope and follow a schematic template in which victim-survivors, perpetrators, and plausible institutional remedies are explicitly identified.

On top of this analytic attention to main stakeholders, it is crucial to consider another relational aspect, which can be derived from themes (hereafter topics) that constitute human rights discourse: drift and span. Topic drift underscores how similar ideas and meanings can be addressed with different vocabularies across time. For instance, “global warming” and “climate change,” while both indicating noticeable changes to Earth’s ecosystems, have slight differences in the scope of issues each term touches on. Recently, another notion, “climate crisis,” has come into use, further highlighting the urgency and severity of environmental issues. These three terms share some semantic core, but nevertheless they are expressed using different vocabularies. Topic span foregrounds how persistent a topic lasts, reflecting that not all topics receive a similar level of long-lasting attention. Topic drift and span can be used as two dimensions, with which a heuristic framework can be constructed.

Combining these dimensions yields four ideal-typical quadrants in Figure 1. Topics characterized by high drift and short spans (HL) are highly volatile. These topics attract intense attention from varying actors only for brief periods.

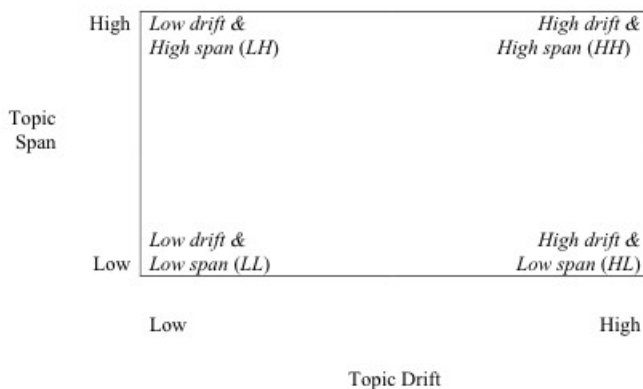


Figure 1. Analytic Frame of Topic Drift and Topic Span

Therefore, they are likely to present rapid changes in vocabularies over time but fail to become embedded as sustained concerns. High-drift topics with high spans (HH) are issues that last longer but the vocabularies used are frequently updated, so their meanings could appear relatively unstable. Low-drift topics exhibit relatively stable semantic characteristics because these topics tend not to present different vocabularies. When such topics emerge only briefly (LL), they likely reflect issues of temporary significance, whereas when they persist over long periods (LH), they are likely to constitute core themes of human rights discourse.

These two relational aspects¹ of human rights discourse—the relationality emerging from multiple co-existing and interacting stakeholders and rooted in how human rights issues are addressed—serve the following hypotheses. First, human rights topics are less likely to be proportionally distributed across the four quadrants. Rather, they are likely to present skewed combinations. For instance, local activists strive to address issues that have not been identified as human rights abuse by formulating new perspectives with new vocabularies, even if their engagement is transient (HL). However, activists also tend to cherish and consistently mention long-lasting issues (LH)—whether they have been properly addressed or not—as their core claims. In other words, low-drift/high-span topics and high-drift/low-span topics are more likely to be present. Topics whose meanings are stable and transient (LL) or topics whose vocabularies rapidly change over a long period of time (HH) would present different prevalence, too. Some enduring issues (e.g., physical and non-physical disciplinary measures in schools or violence in the military) might demonstrate ongoing renewals in vocabulary because similar types of violence often recur in different concrete forms, calling for new vocabulary to describe each instance. In addition, each stakeholder engaging with human rights discourse might show a different tendency in regard to what types of topics they choose to promote. From this discussion, we can develop the following hypotheses:

Hypothesis 1a. Human rights topics are not proportionately distributed across the four quadrants constructed by the degree of topic drift and span.

1 The author appreciates the reviewer who suggested that relationality should be discussed more specifically by distinguishing its behavioral dimension (i.e., how actors select topics) from the semantic dimension (i.e., how topics are related to one another).

Hypothesis 1b. Actors engaging with human rights discourse differ in their selection of topics, distinguished by drift and span.

Though human rights topics can be analytically distinguished and compared by the degree of drift and span, it does not mean that these topics are independent without bridging. As was alluded to above, for instance, activists might frequently connect novel ideas with persistent and familiar narratives and frames (Benford and Snow 2000). Considering that novel ideas do not yet have settled vocabularies, seeing more combinations of high-drift topics with low-drift topics is expected. Alternatively, some human rights issues are closely related to pseudo-legal decisions that result in official remedies may show low-drift topics across different domains, such as in legal vocabularies and the vocabularies persistently promoted by activists.

Hypothesis 2. Topics classified differently by drift and span, nevertheless, are bridged when actors engaging in human rights discourse combine multiple topics.

Also, when multiple topics are bridged, we need to examine the extent to which such bridging is associated with semantic similarity. Intuitively, topics whose meanings are similar are likely to be linked seamlessly, and dissimilar topics are less likely to be bridged. For instance, a human rights text addressing women's rights is likely to contain topics that serve related ideas, such as women's increasing representation in electoral politics or the gender pay gap. However, it is not equally intuitive whether topics that are semantically similar but assigned to different drift-span indexes are still likely to be connected because of their similarity. This point is crucial in understanding how new issues—whose potential longevity is not yet determined—are forged. To identify what does not appear as a human rights violation case on the surface, such novel issues might need to be connected to long-standing, low-drift topics (LH), so that the associated low-drift topics can provide a familiar framework, thus allowing the new issue to be portrayed as something partially familiar and to allude to well-known issues. High-drift topics can also be easily connected with other high-drift topics in an effort to find some common points, even if they are semantically dissimilar, so that they can be synthesized as potentially noteworthy issues. When these topic bridges are made, they may not necessarily be semantically similar.

To investigate what patterns are empirically observed, we use the following hypothesis:

Hypothesis 3. Both semantically similar and dissimilar topics can still be bridged when considering topic drift-span positions.

Data and Methods

Data

This study constructs a new dataset consisting of 39,415 Korean-language documents produced between 1990 and 2010 by three stakeholders—local human rights and civil society organizations, news outlets, and the NHRCK—

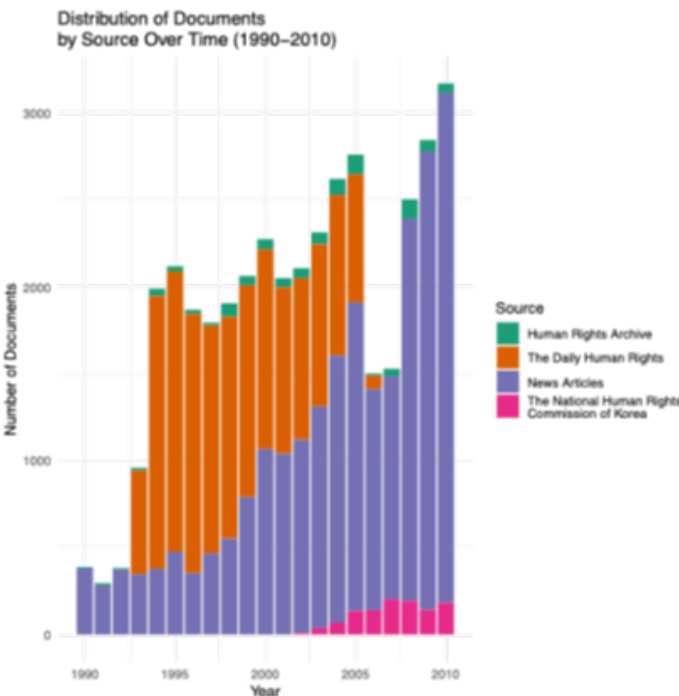


Figure 2. Human Rights Texts from Four Sources in Korea (1990–2010)

Note: This figure shows the distribution of human rights texts by source. Overall, the volume of documents increases over time, with the exception of 2006 and 2007, during which the number of newspaper articles is slightly lower than in preceding and subsequent years.

across four sources.² A primary data collection criterion of selecting a text, regardless of its authorship, is whether it contains the terms “human rights” or “rights.”

The media corpus in this study comprises 22,525 newspaper articles collected from BigKINDS, a database managed by the Korea Press Foundation. For the study, all available news articles are included as long as they include the search words. This inclusive strategy is intended to capture more broadly how news outlets discuss human rights issues through a diverse range of stories. The second corpus draws from the NHRCK website and includes 1,106 publicly available official decisions on petitions submitted for review. Under the National Human Rights Commission of Korea Act, any individual or group, including foreign national residents, may submit a petition alleging human rights violations or discriminatory practices specified under the Korean Constitution, either on their own behalf or on the behalf of others. The NHRCK reviews allegations involving public institutions—such as central and local government offices, detention centers, and protective facilities—as well as certain harms committed by private individuals or organizations. Although the NHRCK’s recommendations are not legally binding, they have generally been respected in practice (Koo 2014). This corpus therefore adds an institutional layer to the dataset by capturing how grievances and abuses are framed as claims seeking remedies through formal channels. Redacted versions of these recommendations are made publicly available without identifying information about the petitioners or alleged violators.

Human rights texts produced by civil society organizations come from two related but independent sources. The first is *Human Rights Daily* (hereafter *HRDaily*), a daily newsletter published by Inkkwon undong saranbang between 1993 and 2006. *HRDaily* comprises approximately 3,000 issues and 14,770 articles over a 14-year period and functioned as an activist-oriented news publication circulated among individuals and organizations engaged in human rights advocacy. The second source is an archive maintained by the Human Rights Research Center, Chang, which contains a wide range of documents produced by civic groups addressing human rights issues. From this archive (hereafter HRarchive), we retrieved 1,104 documents ranging from petitions

2 The author is grateful to Bhargav Dave for his assistance with data collection, made possible by a mini-research assistantship from the Data Analytics and Computational Social Science at UMass Amherst.

and press releases to booklets and pamphlets intended to communicate human rights agendas to the general public as well as to fellow civic organizations.

Though it is essential to incorporate human rights texts produced by the aforementioned sources for the study, this decision admittedly comes with analytic tradeoffs. As Murdie, Davis, and Park (2020) note, compiling an exhaustive list of civic organizations is inherently challenging. This lack of representativeness becomes especially salient for HRarchive, which was not designed to be an exhaustive archive. Also, these materials may overrepresent human rights texts produced by progressive organizations. Moreover, activist movements about feminism and the environment, widely accepted as key players in promoting human rights, have created separate archives that may not be properly represented in the current dataset. Despite these limitations, it remains one of the most comprehensive repositories of human rights–related texts in Korea and thus provides indispensable coverage of activist discourse.

The current makeup of the dataset also introduces another analytical and theoretical challenge that should be clearly mentioned.³ HRarchive and NHRCK documents, while relatively long compared to that of other sources, are far fewer in terms of total document count.⁴ This imbalance warrants a theoretical discussion on what the unit of the analysis should be. When a longer document is weighed equally with a shorter document during analysis, the equality of each author becomes prioritized. Theoretically, this approach can be justified. However, longer documents are known to deteriorate the functionality of computational text analysis tools (see the Methods section). An alternative, widely recommended solution is to split the long documents into smaller chunks. This can enhance the quality of analysis but introduces new issues regarding whether each chunk stemming from the same document can be independent in downstream analysis. These issues and proposed solutions are discussed in the following sections.

3 The author again appreciates the reviewer who recommended discussing this more explicitly.

4 There is a sharp and unexpected decline in the number of documents from media outlets in 2006 and 2007, accompanied by a pronounced drop in documents produced by civic organizations in 2006. Print news coverage during these years is lower than in both the preceding and subsequent periods. Because NGOs and social movement organizations typically produce materials on a regularized cadence, this temporary decline is unlikely to reflect a substantive contraction in human rights–related activity; it is more plausible that it indicates gaps in data collection or archival coverage. As the reviewer suggested, removing those years from the data can resolve this discrepancy. However, for this current study, these two years are included; we leave investigation into the reason behind this sudden drop for future study.

Methods: BERTopic Modeling and Topic Drift-Span

BERTopic modeling

This study applies a dynamic BERTopic model (Grootendorst 2022), which has a modular architecture that allows researchers to flexibly combine various methods at each stage of analysis, including different model embedding, dimensionality reduction techniques, and procedure clustering. BERTopic models outperform earlier topic models in several ways. First, unlike earlier models that do not address word order in a document, BERTopic's incorporation of an embedding model allows the researcher to vectorize a dataset based on a language model pre-trained on other corpora, keeping the sequence of words intact. The current study uses jhgan's Korean-language embedding model, which transforms each document into a 768-dimensional vector.⁵

Second, BERTopic's core task—classifying documents into topics that have similar embedding vectors by capturing shared semantic meaning—is carried out through multiple steps. These vectors are compressed into a lower-dimensional space using Uniform Manifold Approximation and Projection (UMAP) and then are clustered using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), which groups semantically similar documents together. This procedure is followed by the application of class-based Term Frequency-Inverse Document Frequency (c-TF-IDF). This allows the model to specify the importance of words within a topic.

The whole pipeline utilizes a couple of parameters set by researchers, such as the minimum cluster size for HDBSCAN to define the smallest number of documents for a topic. However, unlike earlier topic models, researchers do not set a potential number of topics in advance. Rather, a BERTopic model assigns a topic per document, including outliers that the clustering algorithm fails to assign. This feature, however, becomes less reliable when the model processes longer documents. Following convention, raw documents (hereafter parent documents) were uniformly split into shorter chunks capped at 500 words maximum. For the current study, outliers specified for the first analysis were embedded again using the same model to see if they were mistakenly specified as noise. After this, dynamic topic modeling is performed to track the prevalence

⁵ More information on this is available at <https://huggingface.co/jhgan/ko-sroberta-multitask>.

and shifts in the language surrounding the specified topics.

Topic drift and topic span

As parent documents were all processed to have a similar length (max. 500 words), some can have multiple topics. While this form of splitting helps longer documents be properly represented, it simultaneously imposes a challenge in measuring topic drift and span by treating all 500-word chunks as independent. To address this issue, topic drift that uses c-TF-IDF vectors to capture changes in vocabularies is computed in two steps. First, for each chunk, the c-TF-IDF vector is computed. Then, within each parent, the average of the chunks is taken, so that one vector per parent is satisfied. If a parent has multiple topics, the topic that appears most frequently is weighed and becomes the modal topic. Then the following formula is used:

$$drift_t = 1 - \frac{1}{A_t} \sum_{(y,y_1) \in A} \cos(v_{(t,y)}, v_{(t,y_1)})$$

The formula compared topic 's parent-balanced c-TF-IDF vectors over two consecutive years. If a topic appears in three consecutive year pairs, then the average of those three pairs is subtracted from 1. In other words, the more a topic shares similar vocabularies from year to year, the larger its average grows and the smaller its drift score becomes.

Topic span is computed by counting how many years a topic appears in and then normalizing that with the number of years of the entire corpus. As the data contains documents from 1990 to 2010, all topics share the same value:

$$span_t = \frac{Y_t}{Y_{max} - Y_{min} + 1}$$

where indicates the count of years where topic appears, and denote the last and first year in the data, respectively. To match topic drift, span is also parent-balanced using the modal topic.

As a result, a chunk has the following variables: author (*HRdaily*, *HRarchive*, *NHRCK*, and *news*), year (1990–2010), and topic assignment. The data can also be resorted to different questions. For instance, longer documents can be assigned to multiple topics, and each topic can appear in multiple years, which can be distinguished by using a topic-year index. By design, a topic gets one drift and span quadrant assignment.

Results

Descriptive Summary: The Human Rights Corpus, 1990–2010

Figure 3 presents the distribution of documents in two ways. The blue bars indicate the share of each source, using parent documents as the unit. As discussed earlier, newspaper articles that reference human rights or other rights account for 56.7%, followed by *HRdaily* (37.1%), and this jointly represents about 93% of the entire corpus. However, this distributional composition radically changes with a 500-word cap implemented. In this setting, HRarchive chunks represent 35.5%, followed by chunks from the news (34.7%), *HRdaily* (24.3%), and the NHRCK (5.4%). This noticeable imbalance stems from several reasons, one of which being that the current BigKINDS service does not provide full news articles. Therefore, all newspaper articles are relatively shorter. Second, though HRarchive and *HRdaily* encompass texts generated by activists, they seemingly present their claims and campaigns differently depending on the chosen medium. *HRdaily* is similar to conventional newspapers, targeting civil society actors. Therefore, *HRdaily*'s parent documents are relatively short. In contrast, HRarchive and NHRCK documents are longer, signaling that they often present their claims and decisions at length. To address this imbalance, parent-balancing is implemented whenever necessary.

BERTopic models were fit to 65,051 chunks, which yielded 2,153 topics for the entire corpus. Of the chunks, 21.4% were classified as outliers and

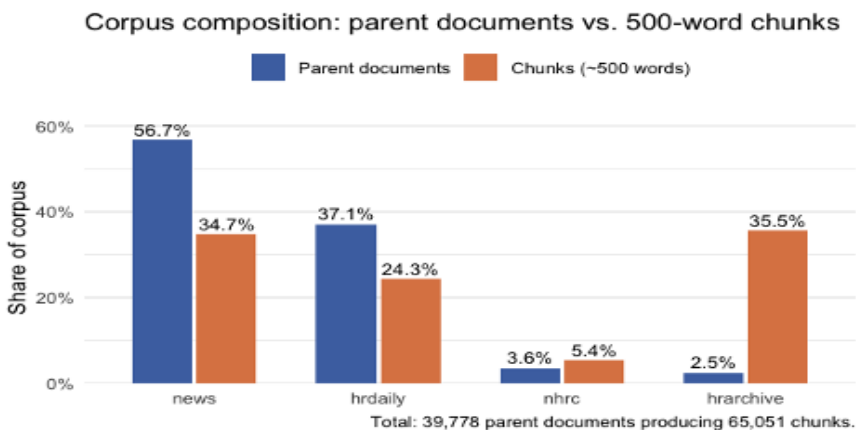


Figure 3. Corpus Composition

Table 1. Parent-level Source and Topics

	Document count	Total topic counts	Topics with measurable drift (%)
HRarchive	787	91	21 (23.1)
<i>HRdaily</i>	10897	681	589 (74.7)
News	16051	1314	884 (67.3)
NHRC	976	67	29 (43.3)
Total	28,711	2,153	1,443

Note: Document count reports the number of topics that has at least one parent document. As a topic cannot be used to compute drift, the number of topics with measurable drift becomes smaller.

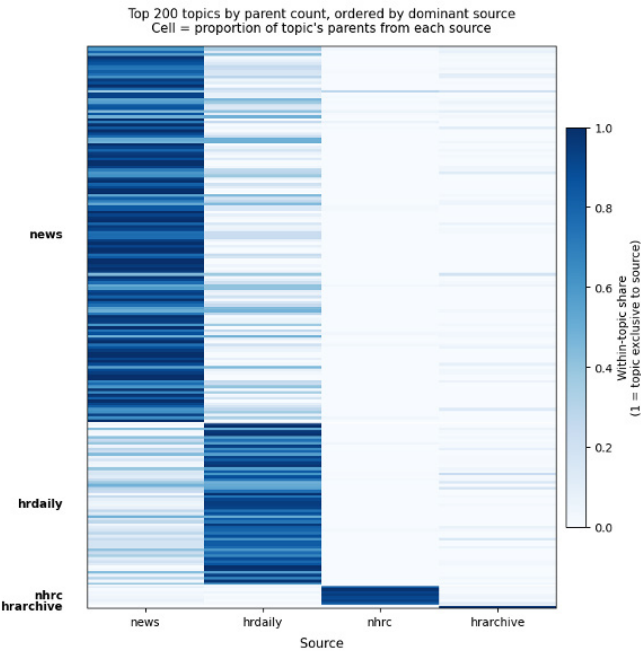


Figure 4. Top 200 Topics by Source Share

removed from the study, accordingly. Single-chunk parents—whose original document has fewer than 500 words—account for 93.5% of the corpus. The average number of chunks per parent is 1.58, and the maximum is 102. Table 1 reports the topic distribution across the corpus with the number of topics that have measurable drift scores. In Table 1, the difference between the total number of topics assigned to a source as its dominant topic and the number of topics with measurable drift shows an interesting distributional change. For instance, whereas about 75% topics assigned to *HRdaily* appear in two consecutive years,

the same can be said for only 23% of HRarchive-associated topics. This implies that HRarchive tends to have more topics that stay active for two years less than other sources.

On top of this distributional characteristic, another interesting feature of the corpus is HRarchive's share of the top 200 topics in Figure 4, which visualizes each topic as a four-column grid representing the relative contributions of the four sources, using the parent document as the unit for counting. Cell shading indicates source prevalence, with the darker shade corresponding to a higher proportion of documents from the respective source. The news and *HRdaily* reveal an unexpectedly high level of overlaps, as many topics are addressed in both at a similar rate. NHRCK documents are also quite interesting because most of the top-ranked topics are rarely mentioned, though HRarchive's contributions seem highly distributed across many topics.

Topic Drift and Span

Figure 5 presents human rights topics specified by a dynamic BERTopic model in a two-dimensional space defined by drift on the X-axis and span on the Y-axis. For this analysis, 1,443 topics (representing 28,711 parent documents, see Table 1) are filtered out from all topics whose drift score cannot be measured (e.g., a topic that appears in only one year). As topic drift and span both are highly skewed in different directions, their corpus-wide medians are used to identify thresholds to split all topics into two parts—low and high—along each dimension: 0.974 for topic drift and 0.1905 for topic span. Given that the upper bound for both topic drift and span is 1, the median of the topic drift is relatively high and that of topic span is relatively small. Coordinated together, the joint distribution by drift and span shows a steep upward curve, and this is also confirmed with the subtotal of all high span topics, which account for 83.7% of the total regardless of their drift level in Table 2.⁶

This empirical joint distribution supports our first hypothesis (H1a) that topics are not proportionally distributed across four quadrants, though the effect size is small ($\chi^2 = 5.933$, $df = 1$, $p = 0.0145$, Cramér's $V = 0.0641$). In other words, the discursive space is highly dominated by topics that tend to stay active longer,

⁶ Spearman's r test shows that topic drift and span are not correlated

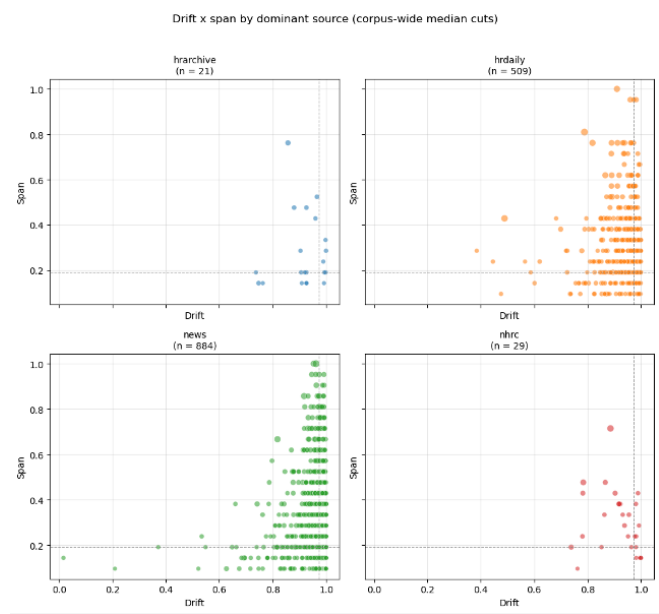


Figure 5. Topic Drift and Topic Span by Source

Note: To construct a two-dimensional drift-span topic space, the median values are used for both (drift median = 0.974, span median = 0.1905).

Table 2. Topic Drift and Topic Span

Drift (%)	Span		Total (%)
	High	Low	
High	622 (81.6)	100 (13.9)	722 (100)
Low	586 (81.28)	135 (18.7)	721 (100)
Total	1208 (83.7)	235 (16.3)	1,443(100)

Note: Each row sums to 100%.

even if their vocabularies are rapidly updated.

This discussion can be furthered by testing whether this pattern appears consistently across sources. With the counts of parent documents assigned to modal topics, Table 3 indeed confirms that all sources share the pattern most prominently with LH and HH topics, which together account for 92.8% of the total parent documents (=1002.66, $df = 9$, $p = 4.61e-210$, Cramér's $V = 0.1079$).⁷

⁷ To check robustness, several additional tests were run with different conditions. In addition to the primary

Table 3. Parent-level Source and Quadrant Contingency Table

	Quadrant Assignment (%)				Total (%)
	LL	LH	HL	HH	
Hrarchive	41 (5.21)	496 (63.02)	15 (1.91)	235 (29.86)	787 (100)
<i>HRdaily</i>	516 (4.73)	7675 (70.43)	174 (1.60)	2532 (23.24)	10897 (100)
News	937 (5.84)	8769 (54.63)	367 (2.29)	5978 (37.24)	16051 (100)
NHRC	3 (0.31)	856 (87.70)	19 (1.95)	98 (10.04)	976 (100)
Total	1497 (5.21)	17796 (61.98)	575 (2.00)	8843 (30.80)	28,711 (100)

Note: Each cell shows the number of parent documents, whose modal topic's quadrant assignment falls in its associated quadrant. Modal topic is used when a parent document has multiple topics. Each row sums to 100%.

More substantively, what is notable here is the prevalence of high-drift and high-span topics—HH (upper right)—in the corpus. Some example topics assigned to HH include Topic 61 on corporal punishment in schools, Topic 21 on lack of varying social support, or Topic 119 regarding illegal wiretapping. With a variety of themes, these topics tend to be frequently discussed over time, and whenever they are, they tend to use different vocabularies. This may reflect substantial challenges and problems that have been identified as recurring representative cases of human rights abuses.

Some examples of LH topics that account for more than half of the entire corpus are as follows: Topic 122 on discrimination faced by people with disabilities, Topic 1492 on children—in particular, special news articles covering events for the annual Children's Day holiday—Topic 1253 on various strikes and protests revealing abuses in the workplace, and Topic 1593 on contention between human rights activists and lawmakers over whether certain bills should or should not be passed. As LH topics tend to be persistent with relatively little change to their vocabularies, it appears that LH topics capture widely established tropes and typified frames.

Another point that warrants discussion pertains to the composition of NHRCK documents (see Table 3). These documents follow the same pattern

test result, which included all parent documents assigned to at least one topic, four additional tests were run by using different samples constructed by (1) a stricter condition (e.g., a topic appearing in at least two parent documents), (2) taking chunks as the unit without parent-balancing, (3) using single-topic parents only, and (4) multiple-topic parents only. The chi-square statistics are all significant and the associated effective sizes are small or moderate: Cramér's $V = 0.123, 0.121, 0.10, \text{ and } 0.19$, respectively.

as the others in terms of LH and HH topics being the most prevalent topics. However, unlike the others, NHRCK documents have more HL topics than LL topics. The NHRCK’s HL topics mostly concern the organization’s recommendations following investigations into individual petitions. Given that there are many topics touching on similar themes, it appears that NHRCK’s formal language used in concrete case descriptions and the short attention given to each case force its topics into the HL quadrant. With these subtle points, we can conclude that the drift-span relationship shows that Korean human rights discourse has many topics that remained active over long periods, regardless of whether they are volatile and demonstrate constant updates in vocabularies.

Bridging Topics

The drift-topic quadrants provide an overview of human rights topics over the course of 21 years. However, they do not reveal whether these identified topics are bridged, and if so, how. As each topic is assigned to a single quadrant, all topics may seem independent, but that is less likely to be the case. In this vein, our Hypothesis 2 addresses this under a specific condition: how the author of a document selects and incorporates multiple topics—i.e., topics potentially assigned to different quadrants—and whether this tendency is statistically significant. To test the hypothesis, we use parent documents that have multiple topics only. According to Table 4, the current corpus has 1,015 parent documents that satisfy this condition. This subsample excludes news articles due to the difficulty in collecting truly comprehensive data.

For this test, we use the PMI test, which is the log-odds of a joint probability of two events to the product of their marginals, to quantify the degree of the co-occurrence of topics assigned to different quadrants. A negative

Table 4. Multi-quadrant Parent Documents by Source

	Total Parents	Single-quadrant Parents	Multi-quadrant Parents	Max Quads	Multi-Quadrant (%)
HRarchive	949	268	681	4	71.8
<i>HRdaily</i>	10,910	10,768	142	3	1.3
News	16,051	16,051	0	1	1.0
NHRCK	1,010	618	192	4	19.0
Total	28,920	27,905	1,015	n/a	100

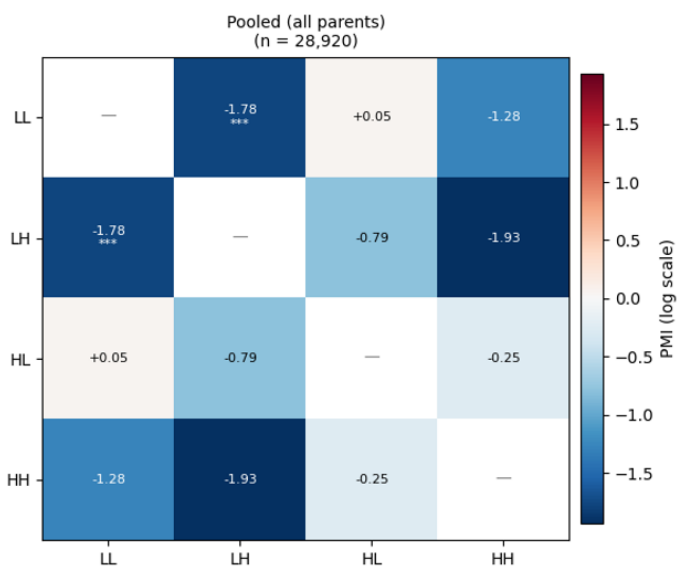


Figure 6. Bridging Topics (Drift-Span Index)

Note: Each cell reports the PMI, with p-values via permutation test.

PMI indicates the two events co-occur less than by chance, whereas a positive PMI signals more co-occurrence and a 0 indicates that the two events are independent. Given that this measure does not provide uncertainty, we added permutation tests with 1,000 repetitions.

The PMIs of the first test that pools up all available parent documents across time is presented in Figure 6. There is only one such case, which is statistically significant, as bridging LL and LH topics is less likely to occur (PMI = -1.78, $p = 0.00$). Table 5 shows that most of topics take place within a single quadrant (e.g., LH-LH, LL-LL, etc.), indicating that fewer documents connect multi-quadrant topics within a single document. The significant negative PMI for the LL-LH implies that authors of human rights texts tend not to connect two low-drift topics (i.e., highly stable) whose degree of span is different (i.e., transient or persistent). This might indicate that once issues have established their vocabulary sets, they are less likely to be mentioned simultaneously.

We can obtain further refined understandings by incorporating two additional points: (1) whether this pattern remains consistent over time and (2) whether different stakeholders share a similar pattern. Figure 7 reports PMI test results applied to period-stratified subsamples by splitting the studied period of time into three parts, each of which is intended to have a similar number of

Table 5. Co-occurrence Topics by Quadrants (Pooled)

	LL	LH	HL	HH
LL	1669			
LH	178	18218		
HL	51	240	839	
HH	153	874	215	9534

Note: Each cell reports the count of parent-documents whose topics appear in a joint or single quadrant. Single-quadrant combinations are not used in PMI computations.

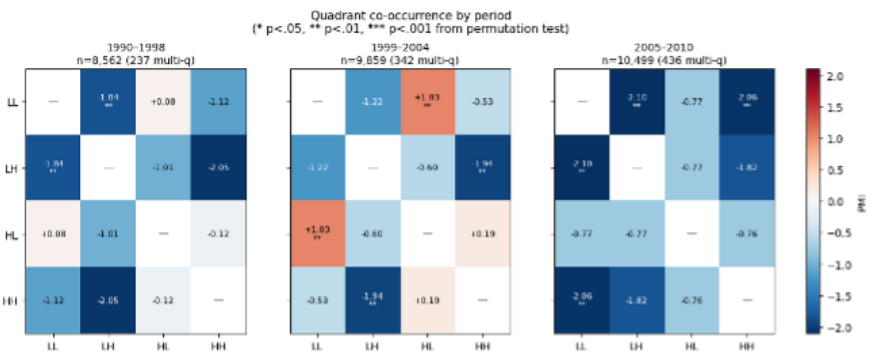


Figure 7. Topic Bridging by Period

Note: Each cell reports the PMI, with p-values via permutation test.

parent documents. LL and LH quadrants are less likely to be bridged in period 1 (1990–1998, PMI = -1.84, $p < .01$) and period 3 (2005–2010, PMI = -2.10, $p < .01$). This result reflects the pooled PMI test result in Table 5. Figure 7 also uncovers some subtle patterns; in period 2 (1999–2004), contributors to human rights discourse bridged LL-HL topics (PMI = 1.03, $p < .01$), whereas in period 3, LL-HH topics were less likely to be connected (PMI = -2.06, $p < .01$). LL-HL, the only bridging with a significant positive PMI, might cover a pair of topics together when both remain transient.

Figure 8 provides clues about how different actors manage a range of topics in a document. Across time periods and topic combinations, HRarchive has PMI = 0, implying that they treat topic pairs independently, with more occurrences of positive bridging between HL and HH in period 2. Given that *HRdaily* is a newsletter circulated among activists, its overarching bridging patterns stand out, with increases in PMIs that bridge LL-LH, LH-HH, and HL-HH, topics even though the PMIs are still negative. This means that, over

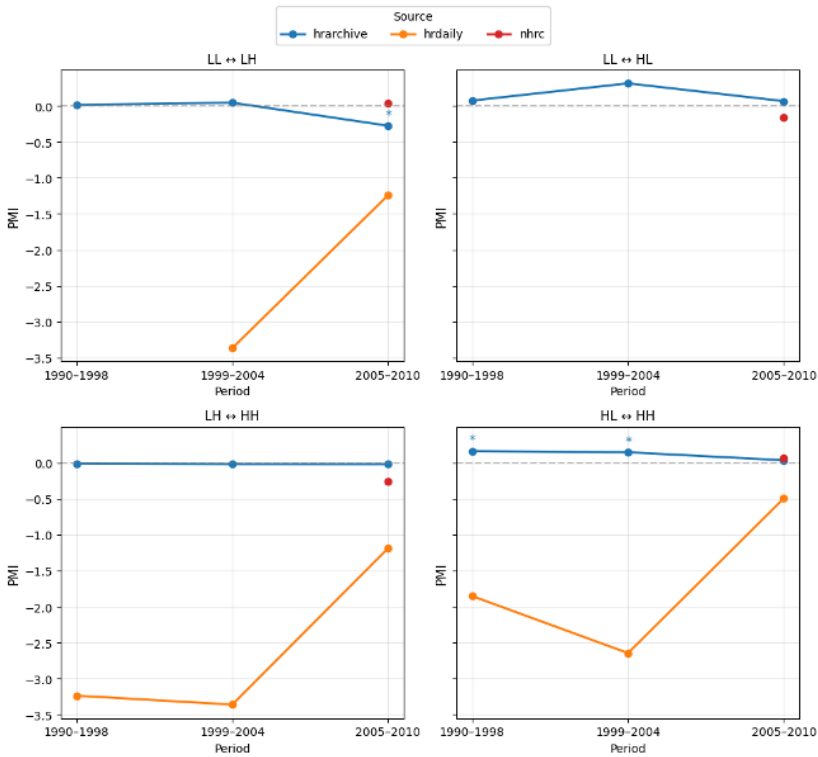


Figure 8. Topic Bridging by Source and Period

Note: This plot is drawn using parent documents that have multiple topics. All news articles are excluded, given that they are all single-topic (or single-quadrant) documents. Partial or no lines indicate no existence of its associated panel case for a source.

time, *HRdaily* articles tend to link these quadrants more than before. Compared to these two activist-centric platforms, the NHRCK shows more consistency in treating topics independently, with PMIs close to 0.

Topic Bridging and Semantic Similarity

Though the PMI tests support our second hypothesis, which predicted distinguishable topic bridging patterns, the question of whether those bridged topics also reflect semantic similarity remains. Attempting to answer this requires us to combine PMIs—i.e., topic selection pattern and semantic similarity between topics. Assuming that dissimilar issues are less likely to be mentioned together (e.g., issues on violence in the military and discrimination against

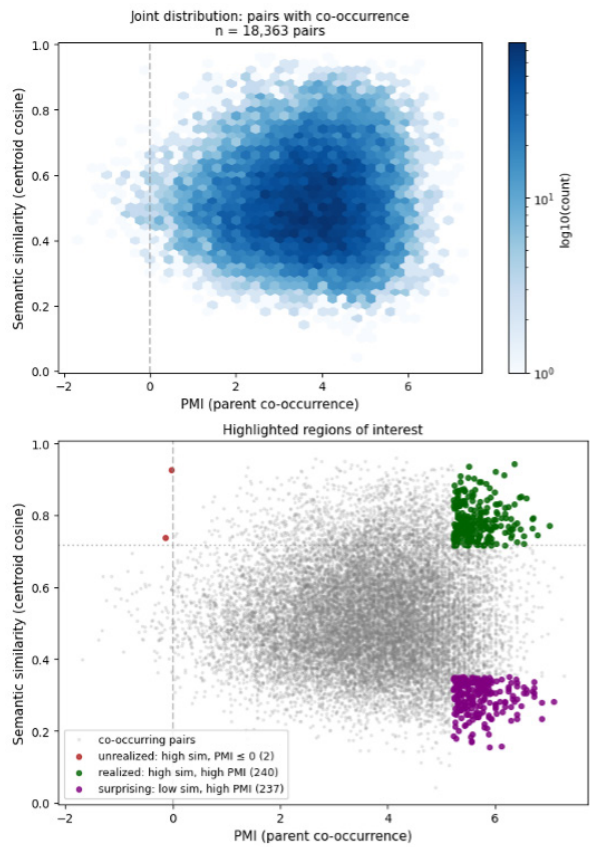


Figure 9. Joint Distribution and Highlighted Regions of Interests

Note: Each hexagon—a pair of topics—is positioned on the plane by semantic similarity on the Y-axis and PMI on the X-axis. To define Realized (green), Unrealized (orange), and Surprising (purple) topic pairs, we set the 95th percentile as the upper bound and the 5th percentile as the lower bound were used.

immigrant workers), the following potential scenarios can be considered. First, we can assess whether semantically similar topics are indeed likely to have high PMIs. We can call these topic pairs Realized. However, semantically similar topics might not actually be realized as a linked topic pair even though that could have been possible. These topic pairs are referred to as Unrealized. Or some topics that are not similar can be connected with a high PMI, referred to as Surprising.

In the top and bottom panels of Figure 9, each hexagon indicates a pair of topics whose position is determined by semantic similarity (Y-axis) and PMI (X-axis). As PMI computation requires a parent document that has at least two

topics, we consider only 18,363 topic pairs that have PMIs out of 1,040,403 possible topic pairs ($1,443 \times 1,442 / 2$). The range of semantic similarity among the considered topics is -0.036 to 0.967.

The top panel illustrates that there is no linear correlation between the semantic similarity and PMI of a pair of topics. Some semantically similar topics are indeed connected, but there are also many that are unconnected. Dissimilar topics are not always set apart, either. Therefore, Hypothesis 3 is supported. The observed independence allows for a high-similarity and low-PMI combination, or low-similarity and high-PMI, and invites further investigation.

The bottom panel, more specifically, underscores the aforementioned Unrealized (orange), Realized (green), and Surprising (purple) cases, using the 95th percentile (upper bound) and the 5th percentile (lower bound) as the thresholds for both semantic similarity and PMI. Two illustrative Unrealized bridges are Topic 93 (HH)–Topic 1158 (LH), with a similarity of 0.93 and PMI -0.03, and Topic 1 (LH)–Topic 27 (LH), with a similarity of 0.74 and PMI of -0.13. The Topic 93–Topic 1158 pair addresses themes related to women’s rights. However, Topic 93 tends to do it by referring to events organized by political groups, activists, or advocacy groups, whereas Topic 1158 treats women’s rights as a part of claims that analytically disclose unfavorable conditions for women workers in relation to childrearing or other insufficient institutional support. In a similar vein, Topic 1 and Topic 27 pertain to the formal and legal decisions on human rights abuse cases from courts or the NHRCK. However, Topic 1 captures news reports that cover such case decisions as news, whereas Topic 27 contains documents that contain concrete case descriptions in a number of NHRCK documents. These two examples imply that similar topics are not always connected.

Table 6 reports the quantity of multi-quadrant combinations that are Realized, Unrealized, and realized as Surprising ties. The counts are based on 100 edges ordered by highest semantic similarity for Realized and Unrealized edges, and 100 edges ranked by low similarity and high PMIs for Surprising edges, using the 95th and 5th percentile thresholds. Notably, there are few Unrealized ties among highly similar topics, suggesting that similar topics tend to be connected. More importantly, Table 6 reveals three noteworthy patterns. First, both Realized and Surprising ties are dominated by associations that contain HH topics: 104 ties—realized or surprising—out of 135 ties include an HH topic. When topics are connected, regardless of their connection type, high-drift

and high-span topics prevail. In other words, HH topics that touch on persistent issues and have vocabularies that are continuously renewed are a key driving force behind topic connection. Second, among all topic pairs by quadrant in Table 6, HH–LH cases are most manifested, implying that core, long-lasting themes with stable vocabularies are well connected with topics that bring in new vocabulary on an ongoing basis. Lastly, transient topics (low-span topics) tend not to be connected to other low-span topics; there are 3 cases of HL–HL and 5 cases of HL–LL.

Conclusion

This study examines the dynamics of human rights discourse in local contexts with BERTopic modeling and social network analysis. The following points are empirically examined and tested. To begin, shaping discourse as a space identified by the drift and span of its constituent topics results in topics being disproportionally distributed and dominated by persistent topics, regardless of whether their vocabularies are trope-like stable or updated continuously (LH–HH). Secondly, when multiple topics are present in a single document, topics that differ in degree of drift and span tend not to be used simultaneously. In particular, *HRDaily* coverage shows a potential increase in connections between relatively long-standing and novel themes, though this trend is not strongly supported. Thirdly, we gain greater insights when this behavioral style of bridging topics with different properties by drift and span is examined in relation to semantic similarity. Similar and dissimilar topics both can be connected, but when this happens, the most salient cases include high-drift and high-span topics as one of the connected topics.

These findings help us understand how human rights discourse has been established. This discourse contains the vocabulary of activists who have worked to advance issues that would not have garnered as much attention without their efforts. It also includes vocabulary that is conventional and stable, allowing for myriad issues to be repeatedly or quickly framed using familiar language. This language, in fact, is linked to formal language, which tends to be less connected and co-occurring with other human rights topics. All results reported in the current study demonstrate that the Korean human rights discourse can largely be described as Realized and Surprising connections in terms of persistent topics,

some of which firmly uphold core themes whereas others continue to evolve linguistically to incorporate both ongoing and emerging issues.

Methodologically, this study showcases the potential of combining data-driven computational methods with theoretically grounded social science inquiries. In particular, this study foregrounds two analytical points. On the one hand, human rights discourse can reveal richer stories when multiple stakeholders are considered simultaneously, thus facilitating investigation into how their similar but distinct perspectives shape human rights themes. This point helps to identify that HROs and media outlets can be sources of new concepts and perspectives from diverse viewpoints, while contributions from the NHRCK subtly deviate from the others with a range of topics that shift often and are transient. On the other hand, there are relational dynamics between low-drift topics—which surface over longer periods of time and consist of core themes—and high-drift topics. How these should be discussed has not yet been settled, even though both are prevalent in human rights discourse.

This study suggests several directions for future research. First, it should be noted that computational text analysis is highly sensitive to corpus construction and representation as well as language models. Though the current study compiled as many broad sources as possible, making generalizations based on these findings should be done with caution. Expanding the dataset to include more diverse human rights–related texts, such as archives on specific social movement branches, would refine our understanding of the discourse. Also, newspaper articles with only partial text available add notable uncertainties. Second, though topic drift and span can specify topic characteristics, the mathematical formulas can be further developed. For instance, the current drift equation only considers topics that appear over consecutive years, ignoring other variations—e.g., topics that are dormant for a while and later resurface. Finally, the counting equation we used for span is relatively simple. Future work could introduce new ways to measure different aspects of temporality, so that the relationship between span and drift can be investigated more thoroughly.

Despite these limitations, the construction of topic drift and span helps us understand human rights discourse as a joint product of multiple stakeholders, who bring to the discussion diverse perspectives and vocabulary. Despite topical diversity, issues they address tend to be long-lasting, sustained by both well-established and evolving vocabulary.

References

- Apodaca, Clair. 2007. "The Whole World Could Be Watching: Human Rights and the Media." *Journal of Human Rights* 6 (2): 147–64. <https://doi.org/10.1080/14754830701334632>.
- Bagozzi, Benjamin E., and Daniel Berliner. 2018. "The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of U.S. State Department Human Rights Reports." *Political Science Research and Methods* 6 (4): 661–77. <https://doi.org/10.1017/psrm.2016.44>.
- Benford, Robert D., and David A. Snow. 2000. "Framing Processes and Social Movements: An Overview and Assessment." *Annual Review of Sociology* 26: 611–39. <https://doi.org/10.1146/annurev.soc.26.1.611>.
- Blau, Judith, and Alberto Moncada. 2005. *Human Rights: Beyond the Liberal Vision*. Lanham, MD: Rowman & Littlefield Publishers.
- Chang, Paul Y. 2015. *Protest Dialectics: State Repression and South Korea's Democracy Movement, 1970–1979*. Stanford, CA: Stanford University Press.
- Chaudhry, Suparna. 2019. "Bridging the Gap: The Relationship between INGO Activism and Human Rights Indicators." *Journal of Human Rights* 18 (1): 111–33. <https://doi.org/10.1080/14754835.2019.1579638>.
- Chua, Lynette J. 2015. "The Vernacular Mobilization of Human Rights in Myanmar's Sexual Orientation and Gender Identity Movement." *Law & Society Review* 49 (2): 299–332. <https://www.jstor.org/stable/43670478>.
- Church, Kenneth Ward, and Patrick Hanks. 1990. "Word Association Norms, Mutual Information, and Lexicography." *Computational Linguistics*, 16 (1): 22–29.
- Evans, Peter B., Dietrich Rueschemeyer, and Theda Skocpol, eds. 1985. *Bringing the State Back In*. Cambridge: Cambridge University Press.
- Fariss, Christopher J. 2014. "Respect for Human Rights Has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108 (2): 297–318. <https://doi.org/10.1017/S0003055414000070>.
- Fariss, Christopher J., Fridolin J. Linder, Zachary M. Jones, Charles D. Crabtree, Megan A. Biek, Ana-Sophia. M. Ross, Taranamol Kaur, and Michael Tsai. 2015. "Human Rights Texts: Converting Human Rights Primary Source Documents into Data." *PLoS ONE* 10 (9). <https://doi.org/10.1371/journal.pone.0137111>.

- org/10.1371/journal.pone.0138935.
- Franklin, James C. 2008. "Shame on You: The Impact of Human Rights Criticism on Political Repression in Latin America." *International Studies Quarterly* 52 (1): 187–211. <https://doi.org/10.1111/j.1468-2478.2007.00496.x>.
- Fuhse, Jan, and Sophie Mützel. 2011. "Tackling Connections, Structure, and Meaning in Networks: Quantitative and Qualitative Methods in Sociological Network Research." *Quality & Quantity* 45: 1067–89. <https://doi.org/10.1007/s11135-011-9492-3>.
- Goodale, Mark. 2007. "Introduction: Locating Rights, Envisioning Law between the Global and the Local." In *The Practice of Human Rights: Tracking Law between the Global and the Local*, edited by Mark Goodale and Sally Engle Merry, 1–38. Cambridge: Cambridge University Press.
- Gordon, Neve, and Nitza Berkovitch. 2007. "Human Rights Discourse in Domestic Settings: How Does It Emerge?" *Political Studies* 55 (1): 243–66. <https://doi.org/10.1111/j.1467-9248.2007.00651.x>.
- Grootendorst, Maarten. 2022. "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." *arXiv*. <http://arxiv.org/abs/2203.05794>.
- Hafner-Burton, Emilie M., and Kiyoteru Tsutsui. 2005. "Human Rights in a Globalizing World: The Paradox of Empty Promises." *American Journal of Sociology* 110 (5): 1373–411. <https://doi.org/10.1086/428442>.
- Heinze, Eric, and Rosa Freedman. 2010. "Public Awareness of Human Rights: Distortions in the Mass Media." *The International Journal of Human Rights* 14 (4): 491–523. <https://doi.org/10.1080/13642980802645804>.
- Hendrix, Cullen S., and Wendy H. Wong. 2013. "When Is the Pen Truly Mighty? Regime Type and the Efficacy of Naming and Shaming in Curbing Human Rights Abuses." *British Journal of Political Science* 43 (3): 651–72. <https://doi.org/10.1017/S0007123412000488>.
- Hwang, Ingu. 2022. *Human Rights and Transnational Democracy in South Korea*. Philadelphia: University of Pennsylvania Press.
- Jung, Minwoo. 2024. "Rights Projects: A Relational Sociology of Rights in Globalization." *Sociological Theory* 42(3): 256–81. <https://doi.org/10.1177/073527512412653>.
- Kang, Jin-Gu 강진구, and Kiseong Lee 이기성. 2019. "Tekseuteu maining eul tonghae bon Jeju Yemen nanmin: neibeo nyus daetgeul eul jungsim euro" 텍스트마이닝(Text Mining)을 통해 본 제주 예멘 난민: 네이버 뉴스

- 맺글을 중심으로 [A Study on Yemeni Refugees in Jeju Island Viewed through Text-Mining-Focusing on Naver News Comment]. *Damunhwa kontencheu yeongu 다문화콘텐츠연구 [Journal of Multi-cultural Contents Studies]* 30: 103–35.
- Karell, Daniel, and Michael Freedman. 2020. “Sociocultural Mechanisms of Conflict: Combining Topic and Stochastic Actor-Oriented Models in an Analysis of Afghanistan, 1979–2001.” *Poetics* 78: 101403. <https://doi.org/10.1016/j.poetic.2019.101403>.
- Keck, Margaret E., and Kathryn Sikkink. 1998. *Activists beyond Borders*. Ithaca, NY: Cornell University Press.
- Kern, Thomas, and Sang-hui Nam. 2011. “Citizen Journalism: The Transformation of the Democratic Media Movement.” In *South Korean Social Movements: From Democracy to Civil Society*, edited by Gi-Wook Shin and Paul Y. Chang, 173–89. New York: Routledge.
- Kim, Cheong-Seok 김정석, Jeongran Kwak 광정란, Miseon Kim 김미선, and Sekyung Park 박세경. 2022. “Tekseuteu maining eul hwaryonghan jangae in dolbom yeongu ui donghyang” 텍스트마이닝을 활용한 장애인 돌봄 연구의 동향 [Text Mining Analysis of Research Trends in Care for Disabled People]. *Teuksu gyoyuk jeoneol: iron gwa silcheon 특수교육저널: 이론과 실천 [Journal of Special Education: Theory and Practice]* 23 (2): 91–122. <https://doi.org/10.19049/jsped.2022.23.2.04>.
- Kim, Jongwoo. 2021. “Hanguk ui pogwaljeok chabyeol geumjibeop ul dulleossan damron jihyeong gwa ijunghwa doen ingwon: pogwaljeok chabyeol geumjibeop yippeop gwajeong ul jungsim uro” 한국의 포괄적 차별금지법을 둘러싼 담론 지형과 이중화된 인권: 포괄적 차별금지법 입법 과정을 중심으로 [Anti-Discrimination Law and Polarized Human Rights: Focusing on the Korean Media Discourse of Anti-Discrimination Law]. *Gyeongje wa sahoe 경제와 사회 [Economy and Society]* 129: 84–117. <https://doi.org/10.18207/criso.2021..129.84>.
- Kim, Sookyung 김수경. 2019. “Sosuja jeongchaek gwa ingwon damron: goyong heogaje wa chabyeol geumjibeop sarye bunseok” 소수자 정책과 인권담론: 고용허가제와 차별금지법 사례 분석 [Policies on Social Minorities and Human Rights Discourse: Focusing on Migrant Workers and Sexual Minorities]. *Sahoe gwahak damron gwa jeongchaek 사회과학 담론과 정책 [Discourse and Policy in Social Science]* 12 (1): 85–110. <https://doi.org/10.22417/dpss.2019.04.12.1.85>.

- Kim, Yooneui. 2023. "A Network Analysis of Naming and Shaming in the Universal Periodic Review." *International Interactions* 49 (3): 287–314. <https://doi.org/10.1080/03050629.2023.2172003>.
- Koliev, Faradj, and James H. Lebovic. 2018. "Selecting for Shame: The Monitoring of Workers' Rights by the International Labour Organization, 1989–2011." *International Studies Quarterly* 62 (2): 437–52. <https://doi.org/10.1093/isq/sqy005>.
- Koo, Jeong-Woo. 2011. "Origins of the National Human Rights Commission of Korea: Global and Domestic Causes." In *South Korean Social Movements: From Democracy to Civil Society*, edited by Gi-Wook Shin and Paul Y. Chang, 77–95. New York: Routledge.
- Koo, Jeong-Woo, and Jaesung Choi. 2019. "Polarized Embrace: South Korean Media Coverage of Human Rights, 1990–2016." *Journal of Human Rights* 18 (4): 455–73. <https://doi.org/10.1080/14754835.2019.1629888>.
- Kwon Kim, Hyun-young, and John (Song Pae) Cho. 2011. "The Korean Gay and Lesbian Movement, 1993–2008: From 'Identity' and 'Community' to 'Human Rights.'" In *South Korean Social Movements: From Democracy to Civil Society*, edited by Gi-Wook Shin and Paul Y. Chang, 206–23. New York: Routledge.
- Lebovic, James H., and Erik Voeten. 2006. "The Politics of Shame: The Condemnation of Country Human Rights Practices in the UNCHR." *International Studies Quarterly* 50 (4): 861–88. <https://doi.org/10.1111/j.1468-2478.2006.00429.x>.
- . 2009. "The Cost of Shame: International Organizations and Foreign Aid in the Punishing of Human Rights Violators." *Journal of Peace Research* 46 (1): 79–97. <https://doi.org/10.1177/0022343308098405>.
- Lee, Namhee. 2011. "From *Minjung* to *Simin*: The Discursive Shift in Korean Social Movements." In *South Korean Social Movements: From Democracy to Civil Society*, edited by Gi-Wook Shin and Paul Y. Chang, 41–57. New York: Routledge.
- McAdam, Doug, and Sidney Tarrow. 2018. "The Political Context of Social Movements." In *The Wiley Blackwell Companion to Social Movements*, edited by David A. Snow, Sarah A. Soule, Hanspeter Kriesi, and Holly J. McCammon, 17–42. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781119168577.ch1>.
- McCammon, Holly J., Courtney Sanders. Muse, Harmony D. Newman,

- and Teresa M. Terrell. 2007. "Movement Framing and Discursive Opportunity Structures: The Political Successes of the U.S. Women's Jury Movements." *American Sociological Review* 72 (5): 725–49. <https://doi.org/10.1177/00031224070720>.
- McLean, Paul. 2017. *Culture in Networks*. Cambridge: Polity.
- Meernik, James, Rosa Aloisi, Marsha Sowell, and Angela Nichols. 2012. "The Impact of Human Rights Organizations on Naming and Shaming Campaigns." *Journal of Conflict Resolution* 56 (2): 233–56. <https://doi.org/10.1177/0022002711431417>.
- Merry, Sally Engle. 2007. "Introduction: States of Violence." In *The Practice of Human Rights: Tracking Law between the Global and the Local*, edited by Mark Goodale and Sally Engle Merry, 41–48. Cambridge: Cambridge University Press.
- Mische, Ann. 2003. "Cross-Talk in Movements: Reconceiving the Culture–Network Link." In *Social Movements and Networks: Relational Approaches to Collective Action*, edited by Mario Diani and Doug McAdam, 258–80. Oxford: Oxford University Press.
- Moon, Seungsook. 2005. *Militarized Modernity and Gendered Citizenship in South Korea*. Durham, NC: Duke University Press.
- . 2024. *Civic Activism in South Korea: The Intertwining of Democracy and Neoliberalism*. New York: Columbia University Press.
- Morgan, Kimberly J., and Ann Shola Orloff. 2017. "Introduction: The Many Hands of the State." In *The Many Hands of the State: Theorizing Political Authority and Social Control*, edited by Kimberly J. Morgan and Ann Shola Orloff, 1–32. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316471586.001>.
- Murdie, Amanda M., and David R. Davis. 2012. "Shaming and Blaming: Using Events Data to Assess the Impact of Human Rights INGOs." *International Studies Quarterly* 56 (1): 1–16. <https://doi.org/10.1111/j.1468-2478.2011.00694.x>.
- Murdie, Amanda M., David R. Davis, and Baekkwon Park. 2020. "Advocacy Output: Automated Coding Documents from Human Rights Organizations." *Journal of Human Rights* 19 (1): 83–98. <https://doi.org/10.1080/14754835.2019.1671173>.
- Murdie, Amanda, and Dursun Peksen. 2014. "The Impact of Human Rights INGO Shaming on Humanitarian Interventions." *The Journal of*

- Politics* 76 (1): 215–28.
<https://doi.org/10.1017/S0022381613001242>.
- Oh, Hye-Sung 오혜성, and Mi-Ae Shin 신미애. 2024. “Ingwon e gwanhan sahoejeok hyeonan bunseok: enron bodo reul jungsim euro” 인권에 관한 사회적 현안 분석: 언론보도를 중심으로 [Analysis of Social Issues on Human Rights: Focusing on Media Reports]. *Hanguk junghap gwahak hoeji* 한국융합과학회지 [Korean Journal of Convergence Science] 13 (3): 143–57. <https://doi.org/10.24826/KSCS.13.3.8>.
- Ovsioitch, Jay S. 1993. “News Coverage of Human Rights.” *Political Research Quarterly* 46 (3): 671–89. <https://doi.org/10.2307/448953>.
- Park, Baekkwon, Amanda Murdie, and David R. Davis. 2019. “The (Co) evolution of Human Rights Advocacy: Understanding Human Rights Issue Emergence over Time.” *Cooperation and Conflict* 54 (3): 313–34. <https://doi.org/10.1177/0010836718808315>.
- Polletta, Francesca. 2006. *It Was Like a Fever: Storytelling in Protest and Politics*. Chicago: The University of Chicago Press.
- Pruce, Joel R., and Alexandra Cosima Budabin. 2016. “Beyond Naming and Shaming: New Modalities of Information Politics in Human Rights.” *Journal of Human Rights* 15 (3): 408–25. <https://doi.org/10.1080/14754835.2016.1153412>.
- Ramos, Howard, James Ron, and Oskar N. T. Thoms. 2007. “Shaping the Northern Media’s Human Rights Coverage, 1986–2000.” *Journal of Peace Research* 44 (4): 385–406. <https://doi.org/10.1177/0022343307078943>.
- Seo, Dong-Jin 서동진. 2005. “Ingwon, simingwon, geurigo seksyueolliti: Hanguk ui seongjeok sosuja undong gwa jeongchihak” 인권, 시민권, 그리고 섹슈얼리티: 한국의 성적 소수자 운동과 정치학 [Human Rights, Citizenship and Sexuality: Sexual Minority Movements and Politics in Korea]. *Gyeongje wa sahoe* 경제와 사회 [Economy and Society] 67: 66–87.
- Seokyo Humanities and Social Sciences Research Lab 서교인문사회연구실. 2017. *Hanguk jinbojeok ingwon undong ui yeoksa e daehan ingwon hwaldongga inteobyu jaryojip: 1993-nyeon buteo 2012-nyeon kkaji ui sigi reul jungsim euro* 한국 ‘진보적 인권운동’의 역사에 대한 인권활동가 인터뷰 자료집: 1993년부터 2012년까지의 시기를 중심으로 [Interviews of Human Rights Activists on Korea’s Progressive Human Rights Movements, 1993–2012]. Seoul: Seokyo Humanities and Social Sciences Research Lab.

- Şerban Rosen, Mihaela, and Diana H. Yoon. 2009. "Bringing Coals to Newcastle? Human Rights, Civil Rights and Social Movements in New York City." *Global Networks* 9 (4): 507–28. <https://doi.org/10.1111/j.1471-0374.2009.00266.x>.
- Shafir, Gershon. 2004. "Citizenship and Human Rights in an Era of Globalization." In *People out of Place: Globalization, Human Rights, and the Citizenship Gap*, edited by Alison Brysk, and Gershon Shafir, 11–28. New York: Routledge.
- Shin, Gi-Wook, and Paul Y. Chang, eds. 2011. *South Korean Social Movements: From Democracy to Civil Society*. New York: Routledge.
- Steinberg, Marc W. 1998. "Tilting the Frame: Considerations on Collective Action Framing From a Discursive Turn." *Theory and Society* 27 (6): 845–72. <https://www.jstor.org/stable/658033>.
- . 1999. "The Talk and Back Talk of Collective Action: A Dialogic Analysis of Repertoires of Discourse among Nineteenth-Century English Cotton Spinners." *The American Journal of Sociology* 105 (3): 736–80. <https://doi.org/10.1086/210359>.
- Tsutsui, Kiyoteru. 2017. "Human Rights and Minority Activism in Japan: Transformation of Movement Actorhood and Local-Global Feedback Loop." *American Journal of Sociology* 122 (4): 1050–1103. <https://doi.org/10.1086/689910>.
- Tsutsui, Kiyoteru, and Jackie Smith. 2019. "Human Rights and Social Movements: From the Boomerang Pattern to a Sandwich Effect." In *The Wiley Blackwell Companion to Social Movements*, edited by David A. Snow, Sarah A. Soule, Hanspeter Kriesi, Holly J. McCammon, 586–601. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781119168577.ch33>.
- Winston, Morton E., and John C. Pollock. 2016. "Human Rights in the News: Balancing New Media Participation with the Authority of Journalists and Human Rights Professionals." *Journal of Human Rights* 15 (3): 307–13. <https://doi.org/10.1080/14754835.2016.1153413>.

Eunkyung SONG (eunkyungsong@umass.edu) is a lecturer in the Data Analytics and Computational Social Science at the University of Massachusetts, Amherst.

Abstract

This study examines how human rights discourse has been established in South Korea from 1990 to 2010, focusing on who talks about what, and how those themes are connected or not over time. By representing a novel corpus incorporating these stakeholders' human rights-related texts as topics based on semantic similarities, this study accounts for Korean human rights discourse as a result of how the stakeholders did or did not connect topics that vary in drift (i.e., how a similar theme is framed via different vocabularies) and span (i.e., how long the topics remain active in focus). Korean human rights discourse is largely characterized by the persistence of topics, regardless of whether their vocabulary is stabilized or continuously evolving. Each stakeholder shows a distinct style of bridging topics with different drift-span indexes over time. Topics with similar levels of drift and span are bridged predominantly, and it is less likely for topics whose drift-span indexes are different to appear together in a single document. However, when semantic similarities are considered, it becomes clearer that enduring topics with low drift consist of core themes, and topics that are similarly persistent with high drift play an important bridging role.

Keywords: human rights discourse, BERTopic model, topic drift, topic span