Framing Identity and Gender in Public Discourse: A Corpus Analysis of Representation of North Korean Female Defectors

Sun-Hee Lee and Beomil Kang

This corpus-based study scrutinizes South Korean media's portrayal of North Korean defectors (NKDs), especially women. Analyzing data from five leading newspapers (*Chosun Ilbo*, *DongA Ilbo*, *JoongAng Ilbo*, *Hankyoreh*, and *Kyunghyang Sinmun*) and *Women News*, this article explores power and gender dynamics in media language shaping NKD identities. Using corpus tools, statistical analysis of keywords, and collocations, the study integrates critical discourse analysis of narratives obtained through concordance search. Key findings reveal the significant underrepresentation and misrepresentation of NKD women, often stereotyping them only as victims of violence. The research advocates for a broader gender perspective in media coverage and diverse portrayals of NKD women and men as integral members of the community, emphasizing the need for gender-sensitive and inclusive language in South Korean media.

Keywords   North Korean defector (NKD), gender, corpus, media, discourse analysis

Introduction

According to the Ministry of Unification in South Korea (2024), as of December 31, 2023, over thirty-four thousand North Korean defectors (NKDs) resided in South Korea, and they are predominantly women (72%). Despite their growing numbers, NKDs face persistent socioeconomic challenges and prejudice. Scholars like Chung (2008) and Kim (2012) have examined both the complex and multilayered process of formulating NKD identities and their in-between position across the divided Koreas by tracing the discursive shifts of social labels in relation to the historical changes of social policy toward North Korean migration. These labels include "defectors," "refugees," "defected North Korean residents," and "North Korean migrants." Lee (2019) extends this analysis, arguing

that NKDs' marginalization and fluid identities result from a mixture of Cold War politics, a divisive system between North Korea and South Korea, divisive language, and cultural politics of neoliberal governmentality that prioritizes the formation of individuals as productive citizens within a capitalist framework.

Recognizing the impact of social labeling, this study explores the interplay between cultural politics and language among various social actors, including NKDs, media, and government. Language functions as a critical factor in establishing identity, status, and power (Fairclough 1989; Wodak 2012). Through critical discourse analysis, Fairclough (1989) demonstrated how language contributes to the establishment and reinforcement of ideologies and power relations. Similarly, Van Dijk (1988) explored news as public discourse, connecting the structures of news discourse with the cognitive processes of news production and understanding. As a pioneer of critical discourse analysis, he highlighted that language in news media is closely intertwined with power and ideology, serving as a tool through which these elements are both reflected and shaped. According to him, the way news is structured, narrated, and presented is not neutral but is instead laden with ideological content that serves to maintain or challenge existing power relations. Hall (1997) highlighted that media functions as an active agent in identity formation within cultural and societal frameworks.

This study focuses on the media language shaping public discourse about NKDs, with an emphasis on female defectors, given the gendered nature of North Korean migration. Over the past decade, rigorous studies have examined the representation of North Korean women defectors in television shows, such as KBS's "Love in Asia" and Channel A's "Now, I Am Coming to Meet You," to analyze the gendered discourse of media that results in otherizing NKD women (Tae and Hwang 2012; Kwon 2013; Lee 2014; Oh 2016; among others). Whereas most studies address qualitative analyses on constructing cultural identities of NKDs through media representations, there remains a notable gap in scrutinizing the comprehensive use of language in public discourse, especially from broader, macro-level perspectives. This study addresses this gap, employing a corpus-based methodology to analyze media representations of NKD women in major South Korean newspapers published in 1994-2023, divided into conservative and liberal categories, and a feminist weekly, *Women News*, as a comparison. A corpus is a large body of naturally occurring language data stored on computers. By combining corpus tools and statistical analysis with critical discourse analysis, this study aims to provide empirical insights into the gender and power dynamics in media representations of NKDs.

This article is structured as follows. The next section establishes the foundation by discussing the background and methodology employed in the construction and analysis of the corpus, along with providing an overview of the general portrayal of NKDs in mainstream media. The article then examines corpus data specific to NKD women, contrasting their depictions in mainstream

news outlets with that in *Women News*, a weekly publication established in 1988 that has a feminist stance. Next, a corpus-based analysis of gender bias in media language is presented through the examination of keywords and collocations. The final section offers the conclusions.

Background and Methodology

In the pursuit of understanding the complex narratives surrounding North Korean defectors, this study utilizes a new, combined approach of corpus-based critical discourse analysis. Previous research has demonstrated that the intersection of critical discourse analysis (CDA) and corpus linguistics (CL) offers a robust framework for this investigation (Baker 2004, 2006; Baker et al. 2008; Hardt-Mautner 1995; among many others). CDA is understood as a problem-oriented interdisciplinary analysis rather than a specific method, focusing on a shared interest in power, ideology, injustice, and domination. Thus, CDA is not tied to particular methods but instead adopts any that are adequate to realize its research goals (Baker et al. 2008). Stubbs (2001) argues that CL plays a role in uncovering patterns of language and evaluative meanings. These patterns are not merely personal or idiosyncratic but are widely shared within a discourse community. Using computational procedures, corpus linguistics helps reduce researcher bias, allowing for the examination of large-scale patterns rather than selectively choosing instances that might support a particular stance. This methodology can reveal systematic tendencies and repeated patterns in discourse, contributing to understanding phenomena such as cultural stereotypes and the hidden power of media discourse. By studying repetitive language patterns, researchers can identify how newspapers subtly influence attitudes over time, which Baker (2006, 13) refers to as the "incremental effect of discourse."

The collaboration between CDA and CL is not just a theoretical alignment; it has profound implications for interpreting and understanding the stories of NKDs. While qualitative insights provide depth and context, capturing the human dimensions of the narrative, quantitative data add a level of objectivity, ensuring that the analysis is grounded in measurable and verifiable facts. The convergence of CL and CDA provides a comprehensive toolkit for analyzing discourse (Cheng 2013; Baker et al. 2008; Baker, Gabrielatos, and McEnery 2013), including the representation of specific groups like NKDs. However, a balanced approach that aligns both qualitative and quantitative analyses, along with careful corpus construction, is essential to achieve meaningful results. This study elicits empirical linguistic patterns and distributions reflecting media attitudes and representations, with a close examination of context using corpus tools. The focus will be on how the media formulate the identity of male and female defectors, thereby perpetuating gender stereotypes and prejudice.

To construct an NKD newspaper corpus, we scraped online articles published from January 1990 to December 2022 in five major newspapers, containing writings on NKDs. Each article was archived as a separate text file after eliminating metadata, advertisements, and other non-relevant content using Python scripts. The NKD corpus comprises two subcorpora with distinct political stances: conservative (*Chosun Ilbo*, *DongA Ilbo*, and *JoongAng Ilbo*) and liberal (*Hankyoreh* and *Kyunghyang Sinmun*).[1] As noted by Koo and Choi (2019), the South Korean media landscape is sharply polarized, especially on security-related issues. Conservative outlets, like *Chosun Ilbo, JoongAng Ilbo*, and *DongA Ilbo*, generally support conservative governments and their political agendas, whereas liberal outlets, such as *Hankyoreh* and *Kyunghyang Sinmun*, are more aligned with liberal governments advocating for expanded freedoms of assembly and association and policies favoring the working class and marginalized groups (Koo and Choi 2019, 468-9). Despite their polarized political perspectives, these media outlets display paradoxical attitudes towards NKDs, reflecting a complex and sometimes contradictory stance influenced by their ties to North Korea. Employing a corpus-based methodology allows for an in-depth and objective investigation of the discourse shaped by these media, identifying repeating linguistic patterns and lexical associations.

To ensure that the corpus contained as many articles which referenced NKDs, seven search queries were used including 탈북 *talbuk*, 탈북자 *talbukja*, 탈북민 *talbukmin*, 새터민 *saetemin*, 북한이주민 *bukhanijumin*, 북한이탈 *bukhanithal,* and 하나원 *hanawon*. Using these queries, 22,446 articles containing 6,683,388 *eojeols* (space-based units corresponding to words) were collected and stored in the NKD corpus. Table 1 shows the number of newspaper articles containing NKD-related words and total eojeols. The published article-counts indicate that conservative newspapers, particularly *Chosun Ilbo*, have been leading the discussion on NKDs.

Table 1 illustrates that *Chosun Ilbo* published more articles on NKDs than *Hangyeoreh* and *Kyunghayng* combined, reflecting its right-wing and anti-North Korean stance. This aligns with findings by Lim and Kim (2015) in terms of stories about NKDs published between 2005 and 2013. The conservative newspapers *Chosun Ilbo* and *JoongAng Ilbo* published significantly more feature stories of NKDs than liberal papers, *Hankyoreh* and *Kyunghyang Sinmun* (*Chosun* 105, *JoongAng* 28, *Hangyoreh* 32, and *Kyunghyang* 15). This simple publication record, however, does not fully provide a complete record of the media discourse on NKDs. Before moving onto corpus-based tools for discourse analysis, we will first examine historical distributions of distinct social labels of NKDs used in the NKD-related news articles. Different social labels referring to NKDs have been identified in previous studies, including those by Chung (2008) and Choi (2016).

Figure 1 depicts the fluctuating labels for North Korean defectors (NKDs) in major South Korean newspapers. The official term since the 1997 Act is 북한이탈

Table 1. Numbers of NKD Articles and Eojeols in the Five Major Newspapers

| Political stance	Newspapers	Number of articles	Number of eojeols
conservative	*Chosun Ilbo*	7,962	2,126,459
	JoongAang Ilbo	4,670	1,425,238
	DongA Ilbo	4,479	1,305,681
	Subtotal	17,111	4,857,378
liberal	*Hangyoreh*	2,747	1,023,533
	Kyunghyang Sinmun	2,591	802,477
	Subtotal	5,338	1,826,010
Total		22,449	6,683,388

Source: Authors

주민 *bukhanitaljumin* "resident escaped from the North." Chung (2008) highlights the negative "escape" connotations of 탈북자 *talbukja*, tracing NKD labels through history—from "system selective migrants" (1945-1950), "war refugees" (1950-1953), and "heroes who returned to the state" (1962-1993), to "new settlers" (2005-2008). Choi (2016) confirms these patterns, noting a shift in newspaper terminology between 2000-2013. Our study differs by focusing on the dispersion of term usage across articles rather than their frequency within the articles.

Figure 1 indicates "defector from the North" (탈북자 *talbukja*) as prevalent, despite its negative connotation associated with "escape" and a derogatory meaning of *-ja*, as noted by Chung (2008) and Choi (2016). Alternatives like "North Korean migrants" (북한이주민 *bukhanijumin*), "North Korean refugees" (북한난민 *bukhannanmin*), and "free migrants" (자유이주민 *jayuijumin*) lack widespread acceptance. Introduced in 2005, the term "new settlers" (새터민 *saetemin*) was opposed by NKDs for undermining their political identity (Park 2007). The term 탈북민 *talbukmin*, a condensed form of "residents escaped from the North" (북한이탈주민 *bukhanitaljumin*), is expanding in use. This term (consisting of 탈북 *talbuk* and 민 *min*) is positively received by NKDs, resonating with terms like "citizen" (시민 *simin*), "nationals" (국민 *gukmin*), and 주민 *jumin* "residents." Since 2019, 탈북민 *talbukmin* has become the predominant term in South Korean media, as Figure 1 illustrates.

Bucholtz and Hall (2005) provide a comprehensive framework to understand how social identities are constructed and negotiated through language within the broader context of sociocultural linguistics. Figure 1 demonstrates that NKD identities are dynamic and continually reshaped through sociocultural practices and interactions in discourse. Choi (2016) emphasizes the positive prosody of 이주 *iju* "migration" over 탈북 *talbuk* "escape from the North," suggesting terms like 북한 이주민 *bukhan ijumin* "North Korean migrants" for official use. However,

Figure 1. Distributions of NKD Articles of Search Queries in Five Major Newspapers
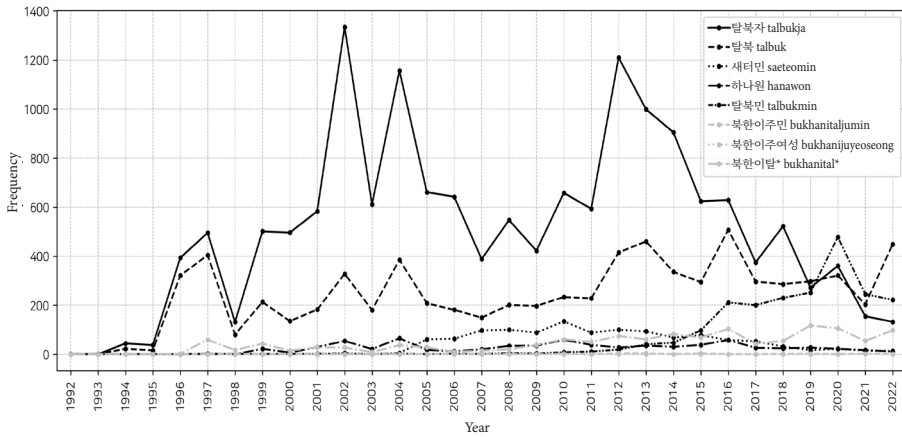
Source: Authors

탈북 *talbuk*-containing terms (탈북자 *talbukja*, 탈북민 *talbukmin*, 북한이탈주민 *bukhanitaljumin*) remain dominant, signifying NKDs' complex sociopolitical position. The fluctuating labeling for NKDs in South Korea indicates that NKD identities are continually evolving, shaped by an interplay between self and society, and rooted in their political status from escaping North Korea. The corpus-based analysis of changing NKD labels across time and regions offers insight into these dynamic identity constructions.

Our analysis of the NKD newspaper corpus involves extracting the top one hundred frequent content nominals, including proper names and common nouns, and categorizing them into eight distinct themes, as in Table 2. This approach is designed to identify the primary topics within the corpus. By excluding some query words, such as 탈북 *talbuk*, time expressions, and grammatically functional bound nouns, the methodology isolates content-driven elements.

The categorized topic indicators in Table 2 reveal a distinct focus of public discourse on political and diplomatic relations, with comparatively less emphasis on community, legal, economic, and educational issues concerning NKDs. This trend in topic distribution highlights a significant leaning towards international and governance-related matters in NKD media discourse. The topic indicators of country/region and politics/governance show three times more token distributions (573,015 versus 172,493) over topic indicators related to domestic, communal, or individual life (community, law/rights, economy, and education). This fact indicates that the South Korean media highlights issues of international and domestic politics in relation to NKDs. Furthermore, the lexical richness in the former categories is more than two times that of the latter categories, reinforcing the observation that NKD media discourse predominantly centers

Table 2. Top One Hundred Indicators of the NKD Corpus (Ranked by Token Frequency)

No	Category	Type	Token	Words
1	country/ region	18	312,538	북한 "North Korea," 중국 "China," 한국 "South Korea," 미국 "U.S.," 북 "North Korea," 남북 "South and North," 서울 "Seoul," 남한 "South Korea," 미 "US," 일본 "Japan," 지역 "region," 평양 "Pyongyang," 국내 "domestic," 세계 "world," 한반도 "country," 한 "Korea," 중 "China," 국제 "international"
2	politics/ governance	37	260,477	정부 "government," 대통령 "President," 통일 "unification," 대북 "North Korea-bound," 김정은 "Kim Jong-un," 관계 "relation," 국가 "nation," 의원 "congressman, deputy," 김정일 "Kim Jong-il," 대표 "representative," 당국 "the authorities," 체제 "system," 국민 "citizen," 장관 "minister," 관계자 "person concerned," 정책 "policy," 유엔 "UN," 난민 "refugee," 위원장 "chairperson," 정권 "regime," 통일부 "department of unification," 외교 "diplomacy," 핵 "nuclear," 대사관 "embassy," 국정원 "NIS," 군 "military," 나라 "country," 당 "political party," 회담 "talk," 경찰 "police," 검찰 "prosecutors," 정치 "politics," 국경 "national border," 망명 "exile," 수용소 "camp," 국회 "congress," 강제 "coercion"
3	issues/ activities/ circumstances	18	119,501	문제 "issue," 사건 "incident," 상황 "situation," 정보 "information," 가능성 "possibility," 과정 "process," 활동 "activity," 조사 "investigation," 이유 "reason," 전단 "leaflet," 입장 "stance," 관심 "interest," 대상 "target," 생각 "thought," 계획 "plan," 결과 "result," 대화 "conversation," 평화 "peace"
4	community/ individuals	10	87,492	사람 "person," 주민 "resident," 사회 "society," 가족 "family," 단체 "group," 출신 "origin," 여성 "female," 아이 "child," 집 "house," 생활 "living"
5	law/rights	3	34,504	인권 "human rights," 자유 "freedom," 법 "law"
6	media	5	27,611	기자 "reporter," 방송 "broadcasting," 영화 "movie," 언론 "media," 사진 "photo"
7	economy/ support	4	25,368	지원 "support," 경제 "economy," 돈 "money," 사업 "business"
8	education	5	25,129	교수 "professor," 교육 "education," 학생 "student," 학교 "school," 대학 "university"

Source: Authors

around political domains.

In Table 2, the prevalence of topic indicators highlighting sociopolitical issues, such as 문제 *munje* "issue/problem," 사건 *sagen* "incident," 인권 *ingwon* "human rights," 출신 *chulsin* "origin," and 대상 *daesang* "target" is notable. The recurring use of terms like 문제 *munje* "issue/problem" and 사건 *sageon* "incident" often indicates conflict-related issues. The frequent mention of 출신

chulsin "origin" suggests a strong association between defectors' origins and their identity representation. The use of 대상 *daesang* "target" indicates a portrayal of defectors lacking agency, often as passive subjects or as groups/individuals. These linguistic patterns not only highlight specific thematic areas but also subtly influence perceptions and attitudes towards NKDs. However, understanding the precise function of these topic indicators require deeper investigation. The presence of these terms alone does not reveal their functions or the narratives they weave within the corpus. For instance, the topic nominal 인권 *ingwon* "human rights" predominantly appears in contexts related to North Korea, rather than NKDs in South Korea. Among 1,066 instances of 인권 *ingwon* "human rights" in conservative newspapers, 422 immediately cooccur with 북한 *bukhan* "North Korea" or 대북 *daebuk* "toward the North." In contrast, only forty-two instances of 인권 *ingwon* "human rights" immediately follow 탈북자 *talbukja* "NKDs." These are mostly part of larger nominal phrases like 탈북자 인권 *talbukja ingwon* "human rights of NKDs," 탈북자 인권 문제 *talbukja ingwon munje* "human rights issues of NKDs," 탈북자 인권 유린 *talbukja ingwon yurin* "human rights violation," and 탈북자 인권 보호 *talbukja ingwon boho* "protection of NKD human rights," among others. Importantly, narratives around 인권 *ingwon* focus on North Korea or China, with few examples highlighting the situation in South Korea, as illustrated in examples (1) and (2).[2]

(1) 조사위원들은 북한 인권 유린을 조사하기 위해 방한했지만, 한국에 정착한 탈북자의 인권 문제도 심각하다고 여긴 듯하다 "The investigators visited South Korea to probe into human rights abuses in North Korea, but they also seem to consider the human rights issues of North Korean defectors settled in South Korea to be serious." (*Chosun Ilbo* August 22, 2013)

(2) 북한인권운동 단체들은 … 북한 주민과 3만 4000명 탈북자의 인권 상황이 오히려 나빠졌다고 호소한다. 그래서인지 어느 정부 때보다 인권운동이 활발하다 "North Korean human rights organizations claim that the human rights conditions for North Korean citizens and the 34,000 defectors have worsened. This may be why human rights activism is now more vigorous than under any previous government." (*JoonAang Ilbo* June 2, 2021)

The corpus examples of 인권 *ingwon* imply the importance of close reading to identify accurate portrayals of these topic indicators. A comprehensive analysis of the linguistic context is essential to discern the interconnected issues. What appears as a simple word can lead to deeper insight into the complexities of the dynamics of the issues. In corpus linguistics, concordance analysis is invaluable in this regard. A concordance displays occurrences of a selected word or phrase within a text, showing the surrounding context on both sides. By sorting these occurrences alphabetically on either the left or right side of the chosen word or cluster, researchers can easily examine different linguistic patterns or nuances, such as semantic meaning and discourse tone. Concordance is not merely

Figure 2. Sample Concordance (KWIC) of 탈북자 "NKD" Using the Software AntConc (v. 4.2)

File	Left Context	Hit	Right Context
조선_2003-...	·라는 생각이 들더라고요." 그래서 명절 때마다	탈북인들을	불러 밥을 같이 먹기 시작한 게 2년 전부터는 '자위
중앙_2004-...	재사관의 한 관계자는 "한국 국제학교에 진입한	탈북자들에	대해서는 중국 당국이 상당한 시간을 두고 검토힐
동아_2003-...	안겨주는 행사도 있었다. 그러나 28명이나 되는	탈북자들은	인천공항에 도착하자마자 국민의 시야에서 사라?
동아_2018-...	든는다는 건 좋죠. 그런데 '태영호를 추방하라', '	탈북	종업원 북송하라'와 같이 위협적인 청원이 올라오
20190405_2...	고도 참을 수밖에 없는 탈북 여성들 2017년 4월	탈북	여성 H 씨는 소개로 만난 남성한테 성추행을 당했
조선_2008-...	루 때(공식적으로 2005년 1월 초부터) 통일부는	탈북자란	말 대신 새터민이라는 용어를 사용한다고 발표했
조선_2000-...	2로 선교활동을 하고 있는 저자의 현장 리포트.	탈북자	꽃제비의 비참한 삶, 종교와 자유를 향한 갈망을틈

Source: Authors

a tool for aligning words within their context; it is a method to understand the underlying patterns and relationships within the text. By examining the concordance lines, a researcher can see how these words function and interact within their natural environment. A further helpful feature of concordances is that the amount of co-text provided can be regulated from a few words on either side of the node to the whole text. Concordance analysis is useful and akin to qualitative analysis (Baker 2008). An example of a concordance, 탈북자 *talbukja*, arranged in the key-word-in-context (KWIC) format can be seen in Figure 2.

In addition to concordance analysis, another useful linguistic tool is collocation. Collocations are words that frequently appear together within a certain distance of one another. Collocation analysis captures the relationship between words and the contexts in which they are frequently used. It involves studying the frequency with which words appear together in a given corpus, highlighting the inherent associations and patterns that might not be apparent through isolated examination. In corpus linguistics, collocations are identified by a specified window span, such as 5L-5R (five words to the left of the word of interest, and five to its right), and statistical measures, such as mutual information, log-likelihood, t-score, and others. While identifying adjacent combinations of neighboring words (n-grams, such as "I think" and "minor changes"), the window approach to collocations of the word "change" may include "notify" or "place," showing broader patterns and associations, as in examples "notify somebody of any changes" and "changes took place." This analysis can expose how language is crafted to articulate specific viewpoints, frame topics, or even enforce stereotypes. As Stubbs (1997) pointed out, certain collocations can influence each other, such as the term "illegal immigrant," which might prompt people to perceive migration from one country to another as inherently wrong. Such observations may reflect broader societal attitudes or discourses, and the identification of these patterns grants researchers the ability to critically assess language and uncover any embedded biases or prejudices. In the NKD corpus, we have identified twenty collocational nominals of 탈북 *talbuk* or 탈북자 *talbukja* with the highest statistical values of log-likelihood, as we see in Table 3.

In the analysis of the topic indicators and collocation of 탈북 *talbuk*, an

Table 3. Top Twenty Nominal Collocates of 탈북(자) *talbuk(ja)* in the NKD Corpus

No	Collocates	Frequency	Co-occurrence frequency	Log-likelihood
1	중국 "China"	39,890	9,716	45,400.61
2	북한 "North Korea"	115,615	12,185	36,245.60
3	한국 "South Korea"	34,067	6,518	26,776.57
4	문제 "problem"	25,132	4,700	18,972.64
5	단체 "organization"	8,554	3,336	18,964.95
6	정부 "government"	30,324	4,412	15,511.91
7	출신 "origin"	7,156	2,627	14,519.31
8	강제 "enforcement"	5,087	2,147	12,598.23
9	지원 "support"	10,506	2,362	10,394.34
10	북송 "repatriation"	3,275	1,576	9,775.98
11	인권 "human rights"	24,018	3,019	9,683.54
12	여성 "female"	6,826	1,960	9,681.99
13	국내 "domestic"	5,679	1,764	9,028.84
14	청소년 "youth"	3,305	1,437	8,528.13
15	난민 "refugee"	6,183	1,718	8,352.89
16	정착 "settlement"	3,082	1,340	7,950.66
17	가족 "family"	9,376	1,857	7,655.32
18	주민 "resident"	15,057	2,064	6,951.85
19	남한 "South Korea"	9,068	1,670	6,624.62
20	공관 "official residence"	1,959	979	6,165.13

Source: Authors

intriguing emphasis emerges on the term 여성 *yeoseong* "female." This focus is further highlighted by the frequent occurrence of bigram sequences, such as 탈북 여성 *talbuk yeoseong* and 여성 탈북자 *yeoseong talbukca* "female NKD," with 244 and 39 occurrences, respectively. In contrast, the equivalent terms for "male NKD," 탈북 남성 *talbuk namseong* and 남성 탈북자 *namseong talbukja*, are found only eight and four times, respectively. This significant difference in frequency suggests a potential underlying gender bias in the discourse, with a tendency to mark female defectors over their male counterparts. This discrepancy may be connected to the fact that female NKDs have both significantly outnumbered male NKDs since 2006 and accounted for over seventy percent of the defector population in recent years. It is commonly observed in sociolinguistic practices, however, that males are typically the unmarked norm, while females are often

distinctly marked (Lakoff 1973; Andrew 1990). These linguistic usages call for a more thorough investigation to uncover the social and cultural forces that shape gender-based identities and representations of defectors. In the next section, we will extract sub-corpora, focusing on women from the NKD corpus, to explore how female defectors are portrayed in mainstream public discourse. To identify gender biases and related issues, we will compare narratives in the mainstream media with those in feminist media through detailed corpus analysis.

Contrasting Narratives: Representation of Women in the Mainstream versus Feminist Media

In pursuit of a deeper understanding of gender bias and its manifestations within media discourse, our study endeavors to juxtapose the linguistic landscapes of mainstream newspapers with those of *Women News* (여성 신문), a specialized newspaper with a distinctly feminist perspective. Our comparative analysis goes beyond a simple examination of lexical choices and thematic occurrences; it explores the intricate ways in which narratives are constructed, gender issues are highlighted or obscured, and societal norms are both questioned and upheld in the discourse on NKDs. *Women News* serves as a crucial comparative because it not only focuses on stories that may be overlooked in mainstream media but it also offers a reinterpretation of events from a frequently ignored female perspective. By analyzing corpora from both mainstream newspapers and a specialized news outlet, we aim to reveal how South Korean media depict the gendered aspects of NKD migration. This investigation into the gender dynamics of media is vital for pinpointing disparities, biases, and opportunities for a more inclusive and equitable representation of NKDs in public discourse.

To examine the portrayal of women defectors in mainstream media, news articles from the NKD corpus were selected using query words 여성 *yeseng* and 여자 *yeja*, which refer to "woman." Additionally, articles from *Women News* were collected using the same seven query words, including 탈북 *talbuk*, 탈북자 *talbukja*, and five others. Due to its relatively recent establishment and weekly publication schedule, the *Women News* corpus is comparatively smaller and includes articles published in 2009-2022.

Table 4 presents the number of articles and eojeols from both mainstream newspapers and *Women News*, along with the percentage of these articles and eojeols relative to the entire NKD corpus.[3] The data shows that the extracted subcorpora of NKD women comprise approximately 17% of the articles and 22% of the eojeols in the original NKD corpus. Interestingly, liberal newspapers tend to feature more content about women defectors, with about 19% of their articles and 26% of their eojeols dedicated to the topic. In contrast, conservative newspapers allocate slightly less coverage, with 16% of articles and 20% of eojeols

Table 4. Numbers of NKD Women Articles and Eojeols

Category	Newspapers	Number of articles	Number of eojeols
conservative	*Chosun Ilbo*	1,257 (16%)	413,148 (19%)
	JoongAang Ilbo	746 (16%)	283,529 (20%)
	DongA Ilbo	749 (17%)	284,430 (22%)
	Subtotal	2,752 (16%)	981,107 (20%)
liberal	*Hangyoreh*	533 (19%)	284,468 (28%)
	Kyunghyang Sinmun	455 (18%)	198,477 (25%)
	Subtotal	988 (19%)	482,945 (26%)
Five newspapers	Total	3,740 (17%)	1,464,052 (22%)
Women News	Total	418	188,481

Source: Authors

focused on womanhood. This data stands in contrast to our previous findings and those of Lim and Kim (2015), which indicated a more active coverage of NKDs in conservative outlets (such as *Chosun Ilbo*, *JoongAng Ilbo*, and *DongA Ilbo*) compared to liberal ones. Although these conservative media sources may offer broader coverage of NKDs overall, our current analysis indicates that NKD women receive less prominence in these outlets than in liberal newspapers.

Table 5 shows the distributions of the top one hundred content nominals in the *Women News* corpus, categorized into eight distinct groups. It is notable that this distribution markedly differs from both the NKD corpus shown in Table 2 and the NKD women corpus.[4] In the *Women News* corpus, the categories of topic indicators with highest frequency, both in type and token,[5] are issues/activities/circumstances and community/individual. This is unlike those in the mainstream newspapers in which the indicators are country/region and politics/governance. This distributional difference indicates that *Women News* emphasizes topics concerning communal and individual life experiences of NKD women in South Korea, spanning from personal and community matters to broader socio-economic and political themes. In contrast, the mainstream newspapers highlight political and international relations and affairs related to NKDs and NKD women.

Topic indicators in Table 5 include words of conflict with negative semantic prosody, particularly those related to gender. According to Louw (1993), semantic prosody or discourse prosody, refers to the connotative meanings words acquire based on their frequent usage in certain contexts, which can be positive, negative, or neutral. This concept highlights how words can subtly influence perceptions and interpretations beyond their direct definitions. The negative topic indicators in Table 5—such as 차별 *chabyeol* "discrimination" and 폭력 *poklyeok* "violence"—

Table 5. Top One Hundred Indicators of the *Women News* Corpus (Ranked by Token Frequency)

No.	Category	Type	Token	Words
1	issues/ activities/ circumstances	29	6,055	문제 "problem," 평화 "peace," 통일 "unification," 성 "sex/ gender," 말 "speech," 활동 "activity," 과정 "process," 이야기 "story," 성폭력 "sexual violence," 프로그램 "program," 관심 "interest," 이탈 "secession," 사건 "incident," 상황 "situation," 차별 "bias," 역할 "role," 목소리 "voice," 대상 "target," 운동 "campaign," 이유 "reason," 분야 "area," 계획 "plan," 시간 "time," 힘 "power," 폭력 "violence," 결과 "result," 포럼 "forum," 역사 "history," 현실 "reality"
2	community/ individuals	24	5,187	사람 "person," 사회 "society," 청소년 "youth," 삶 "life," 자신 "self," 가족 "family," 아이 "child," 주민 "resident," 남성 "male," 문화 "culture," 집 "house," 어머니 "mother," 출신 "origin," 남자 "man," 장애인 "disabled person," 원장 "director/principal," 작가 "writer," 가정 "home," 생활 "life," 세상 "world," 친구 "friend," 개인 "individual," 남편 "husband," 사랑 "love"
3	country/ region	13	4,472	북한 "North Korea," 한국 "South Korea," 서울 "Seoul," 세계 "world," 중국 "China," 남한 "South," 미국 "US," 남북 "North and South," 지역 "region," 대한민국 "ROK," 국제 "international," 일본 "Japan," 국내 "domestic"
4	politics/ governance	18	3,848	정부 "government," 의원 "lawmaker," 단체 "organization," 정책 "policy," 대표 "representative," 국가 "nation," 센터 "center," 국민 "citizen," 국회 "congress," 정치 "politics," 유엔 "UN," 회장 "president," 장관 "minister," 대통령 "president," 위원장 "chair," 난민 "refugee," 후보 "candidate," 이사장 "chair"
5	law/rights	4	1,265	인권 "human rights," 평등 "equality," 법 "law," 피해자 "victim"
6	education	5	1,204	교육 "education, 학교 "school," 학생 "student," 교수 "professor," 학습 "learning"
7	media	4	750	기자 "reporter," 영화 "movie," 사진 "photo," 감독 "director"
8	economy/ support	3	692	지원 "support," 사업 "business," 일자리 "job"

Source: Authors

when combined with 성 *seong* "sex/gender," form compound nouns like 성차별 *seongchabyeol* "gender discrimination" and 성폭력 *seongpoklyeok* "sexual violence." This lexical pairing emphasizes gender-specific challenges faced by defector women, while influencing interpretations of NKDs. In contrast, the words 차별 *chabyeol*, 폭력 *poklyeok,* and 성 *seong* are conspicuously absent from the top one hundred topic indicators in the mainstream media coverage of NKDs or NDK women. This discrepancy seems to be attributed to *Women News*' feminist perspective, which focuses more on gender issues, including inequality and violence, and often portrays women as victims. To deepen the understanding

of this perspective, it is worthy to examine the usage of 피해자 *pihaeja* "victim," which is a collocate of 성폭력 *seongpoklyeok* "sexual violence" and 성폭행 *seongpokhyaeng* "rape." This perspective frequently co-occurs with these words and a topic indicator categorized as law/rights in Table 5.

> (3) 한국인 브로커에 성폭행 당한 한 탈북 소녀의 이야기를 계기로…정부는 현재 피해자 B양의 진술과 주변 인물의 증언 등을 검찰 측에 전달한 것을 비롯해… 필요한 조처를 하고 있다 "The story of a North Korean girl who was sexually assaulted by a South Korean broker prompted… [t]he government [to take] the necessary steps to… include the victim Ms. B's statement and the testimony of others to the prosecution." (*Women News* January 13, 2009)
>
> (4) 또 가족 구성원들을 위한 맞춤형 지원과 여성과 아동이 안전한 사회를 구축하고 장애 여성, 성폭력 피해자, 일본군 위안부 피해자, 탈북 여성 등 사회 소외계층 지원도 강화할 것이라고 말했다 "(They) cited as priority projects a safe society for women and children and strengthening support for vulnerable marginalized groups such as women with disabilities, victims of sexual assault, Japanese comfort women victims, and North Korean defectors." (*Women News* February 18, 2018)

The employment of 피해자 *pihaeja* "victim" in the *Women News* corpus provides an interesting representation of female defectors. In example (3), this term specifically refers to a female defector, highlighting her individual experience. In contrast, example (4) broadens this context, linking NKD women with other marginalized groups like women with disabilities, victims of sexual violence, and Japanese military comfort women. The predominant usages of 피해자 *pihaeja* "victim" align more closely with the example depicted in (4). Out of seventy-eight randomly selected instances of 피해자 *pihaeja*, merely eight were directly linked to defector women, while the majority of cases grouped NKD women with other marginalized females. This grouping of NKD women with other marginalized females underscores shared challenges, but it might also inadvertently perpetuate negative stereotypes, emphasizing their victimhood and potentially adding stigma to their identities. Moreover, topic indicators like 출신 *chulsin* "origin" and 대상 *daesang* "target" are featured prominently in Table 5, mirroring their usage in mainstream media. These terms focus on the NKD women's origins and often portray them in an objectified manner. Terms like 현실 *hyeonsil* "reality" and 일자리 *iljali* "job positions" are also notable, highlighting the economic challenges NKD women face in securing financial opportunities within a capitalist society.

The analysis of topic indicators reveals that, compared to mainstream media, *Women News* provides a more in-depth exploration of the daily lives and communal struggles faced by NKD women within South Korean society, as evidenced by higher frequencies of tokens associated with individual and communal concerns in Table 5. Nonetheless, the language of *Women News* tends to locate NKD women within narratives of sexual violence or human trafficking,

grouping them with other marginalized women in South Korean society. Although such representations highlight the necessity for various supports, they create the potential of overlooking their agency as independent individuals.

Unveiling Gender Bias through Language Analysis: Keywords and Collocation

This section will provide a detailed comparison of mainstream media and *Women News* through critical analysis of keyword and collocation combined with concordance search. To conduct a comparative analysis between the two sets of newspapers, we utilized the corpus-based method of keyword analysis. This involved assessing word frequencies within the conservative and liberal subsections of the NKD women corpus. Our goal was to identify words that appear more frequently than statistically expected in one part of the corpus compared to another. This process employed Python programming for extracting word frequency, keyword and collocation extraction, and the statistical measurement of log-likelihood.[6]

The concept of keyness, closely related to aboutness, refers to the main concepts, topics, or attitudes in the discourse (Phillips 1989, 7-10). Statistically driven keywords spotlight words that occur more frequently in the target corpus compared to a reference, exceeding random chance. This method enables a statistically grounded comparison between different sections of the texts by directing researchers to remove their own bias and engage in a complex linguistic analysis of these patterns to discern ideological attitudes and perspectives (Baker 2004, 348). To illustrate the distinct stances of conservative and liberal media, Table 6 presents the fifty most significant nominal keywords from the conservative newspapers in comparison to the liberal newspapers, ranked by the log-likelihood statistic. Recognizing non-lexical or less relevant lexical keywords that indicate general differences between the writing styles of the newspapers, unnecessary elements were filtered out through concordance examination in our keyword list.

In Table 6, the keyword lists from conservative newspapers reveal their dominant focus on international and domestic politics, emphasizing national security concerns related to North Korea. Featured are countries like China, Thailand, the UK, and France, as well as political figures, such as the North Korean leaders, *Kim Jong-un* and *Kim Jong-il*, and *Jang Seong-taek* (a high-ranking North Korean official) and his wife, *Kim Kyeong-hee*. Notably, conservative papers frequently use terms like 탈북 *talbuk* "escape" and 탈북자 *talbukja* "NKD," reflecting their strong stance on national security and opposition to North Korea. This orientation is further underscored by keywords like 정치범 *jeongchibeom* "political prisoner," 암살 *amsal* "assassination," 강제 *gangje* "enforcement," 수용소

Table 6. Top Fifty Nominal Keyword Lists of Conservative versus Liberal Newspapers (Ranked by Log-Likelihood Scores)

Category	Top fifty nominal keywords
conservative (*Chosun Ilbo*, *JoongAng Ilbo*, *DongA Ilbo*)	북한 "NK," 김정은 "Kim Jong-un," 탈북자 "NKD' 김정일 "Kim Jong-il," 중국 "China," 평양 "Pyongyang," 마약 "drug," 김정남 "Kim Jong-nam," 태국 "Thailand," 장성택 "Jang Sung-taek' 통일 "unification," 김일성 "Kim Il-sung," 이설주 "Ri Sol- ju," 수용소 "camp," 탈북 "escape from NK," 응원단 "cheerleaders," 이만갑 "Imangap TV show," 간부 "officer' ,공작원 "spy," 형사 "cop," 주민 "resident," 풍선 "baloon," 요덕 "Yoduk," 김경희 "Kim Kyeong-hee," 정치범 "political prisoner," 대표 "representative," 암살 "assassination," 강제 "enforcement," 영국'UK," 소식통 "news source," 김여정 "Kim Yeo-jeong," 노동당 "NK Labor party," 원정화 "Won Jeong-hwa," 김한솔 "Kim Hansol," 외화 "foreign currency," 근로자 "worker," 보육원 "nursery," 국경 "national border," 파견 "dispatch," 쌀 "rice," 숄티 "Scholte," 도발 "provocation," 필로폰 "Philopon," 아들 "son," 천안함 "Cheonan Ship," 뇌물 "bribery," 프랑스 "France"
liberal (*Hangyoreh*, *Kyunghyang Sinmun*)	차별 "discrimination," 이주민 "migrant," 간첩 "spy," 북쪽 "the North," 사건 "incident," 국정원 "NIS," 이야기 "story," 작가 "writer," 장애인 "people with disabilities," 유우성 "Yoo Woo-seong," 이주노동자 "migrant worker," 소수자 "minority," 이탈 "escape," 판결 "verdict," 폭력 "violence," 삶 "life," 비정규직 "temporary position," 국가보안법 "national security law," 재심 "retrial," 확성기 "loudspeakers," 자본 "capital," 부양의무자 "support obligator," 타이 "Thailand," 혐오 "hatred," 금리 "interest rate," 한국전쟁 "Korean War," 조처 "action," 활동가 "activist," 집단 "group," 죽음 "death," 일상 "daily life," 보수 "conservatives," 신자유주의 "Neoliberalism," 타인 "others," 페미니즘 "feminism," 전복 "overturn," 남쪽 "the South," 최저임금 "minimum wage," 종업원 "employee," 엔지오 "NGO," 조작 "fabrication," 세월호 "the Sewol ferry," 소수 "minority," 극우 "far right," 장애 "diability," 평화 "peace," 권리 "right," 연명 "survival' , 손전화 "cell phone," 사회 "society"

Source: Authors

suyongso "camp," 도발 *dobal* "provocation," 마약 *mayag* "drug," and 필로폰 *pilopon* "methamphetamine." Additionally, the list includes 천안함 *cheonanham* "ROKS Cheonan," referring to a ship that sunk in a North Korean attack in 2010, and 원정화 *Won Jeong-hwa*, a female NKD who was sentenced for espionage in 2008, which further illustrate their anti-North Korea focus.

In contrast to conservative news, the keyword list of the liberal news suggests a more pronounced focus on domestic issues, encompassing both politics and policy criticism against the right-wing party and conservative governments. This includes keywords 보수 *bosu* "conservative," 국정원 *gukjeongwon* "National Institute of Security," 국가보안법 *gukgaboanbeop* "national security law," 조작 *jojak* "fabrication," 극우 *geugu* "far right," 신자유주의 *sinjayujuui* "neoliberalism." Additionally, the social issues related to minorities and feminism are prevalent, as indicated by keywords like 장애인 *jangaein* "people with disabilities," 소수/소수자 *sosu/sosuja* "minority," 이주민 *ijumin* "migrant," 이주 노동자 *iju nodongja* "migrant

workers," 혐오 *hyeomo* "hatred," and 페미니즘 *peminiseum* "feminism." The keyword list from liberal outlets also includes terms linked to political incidents and scandals, such as 세월호 *seweolho* "the Sewol ferry (disaster)," 유우성 *Yoo Woo-sung* "Yoo Woo-sung" (a high-profile defector erroneously accused of being a spy), and 확성기 *hwakseonggi* "loudspeaker" (associated with the contentious broadcasts between South and North Korea). Furthermore, terms like 북쪽 *bukjjok* "the North" and 남쪽 *namjjok* "the South" are used to denote North and South Korea, respectively, subtly downplaying the political divide by focusing on geographical distinctions. The adoption of North Korean vocabulary, such as 손전화 *sonjeonhwa* "cellphone," in their reporting is noteworthy because it indicates a more positive engagement with North Korean perspectives.

The results of the keyword analysis distinctly illustrate the thematic and linguistic divergences between conservative and liberal news outlets in their coverage of both international relations and domestic issues. Specifically, conservative outlets predominantly focus on international affairs, particularly those concerning North Korea, China, and Thailand, with a discernible emphasis on topics like political prisoners, drug issues, forced labor camps, and acts of provocation, thereby projecting a critical stance towards North Korea. Conversely, liberal outlets are characterized by a more domestic-centric approach, as evidenced by the prevalence of keywords such as "conservatives," "national security laws," "extreme right-wing," alongside references to domestic events and issues, like the Sewol ferry tragedy, North Korean spy scandals, societal hatred, and feminism. Notably, the liberal media's incorporation of North Korean lexical terms suggests a more nuanced, if not amiable, perspective towards North Korea. These findings highlight the significant impact of ideological leanings on media representations, thereby shaping public discourse around crucial international and domestic matters. The divergent thematic preferences identified in this keyword analysis of South Korean media are consistent with the ideological divisions previously addressed by Choi (2010) and Koo and Choi (2019) that reinforce the persistent dichotomy within South Korean politics and media landscapes. Whereas Choi (2010) focused on the stances of politicians and humanitarian activists, with conservatives underlining the rights of North Korean migrants, critiquing the governments of China and North Korea, and often reviving anti-communist rhetoric while advocating for international intervention, leftists were noted for their critiques of US pressure on North Korea. They advocated for engagement rather than isolation to promote peace on the Korean Peninsula and emphasized humanitarian aid for North Koreans (ibid., 223). Similarly, Koo and Choi (2019) observed that during the period of 1990-2016, progressive outlets (*Hangyoreh* and *Kyunghyang*) provided broader discussions on human rights, particularly civil rights and socio-economic rights, compared to their conservative counterparts (*Chosun Ilbo* and *JoongAng Ilbo*). Mirroring the findings of previous studies, our keyword analysis reveals that

conservative news focuses on international norms, while liberal outlets cover a wider array of topics. This suggests that liberal news provides a deeper narrative integration of NKD women with various marginalized groups in South Korea, including women (notably comfort women and victims of sexual violence), individuals with disabilities, migrants, and workers. However, Baker (2004, 352) cautions that keyword lists, if viewed in isolation, may not fully convey their intended insights and could be misinterpreted when stripped of their context. Thus, it is imperative not to draw simplistic conclusions from these lists, such as inferring that liberal news equates the recognition of NKDs with other minority groups in South Korea, solely based on the keywords. For example, 페미니즘 *peminiseum* "feminism" appears as a prominent keyword in liberal news with fifty-six occurrences, yet it never specifically addresses feminism concerning North Korean defector (NKD) women, instead focusing on South Korean women or women in general. In this analysis, we examine the usage of words in contexts by extracting concordances.

In the given lists of keywords, it is notable that neither conservative nor liberal media prominently feature keywords specifically pertaining to the lives of NKD women, aside from 페미니즘 *peminiseum* "feminism." This observation suggests that, despite the broad thematic coverage, the unique experiences and challenges of NKDs as South Korean citizens may not be focal points in mainstream news narratives, indicating a potential gap between media representations and the realities of NKDs' lives. However, if both news corpora retain similar frequencies of NKD women-related words, these words may not appear in their keyword lists. To obtain a more accurate understanding of NKD women's representations in the media, we compare mainstream news with a specialized corpus of *Women News*. Table 7 displays the keyword lists from both the mainstream NKD women corpus and the *Women News* corpus, laying the groundwork for our comparative analysis of the mainstream newspapers and *Women News*.

The top fifty keywords of mainstream news versus *Women News*, ranked by log-likelihood scores, are presented in Table 7 and exhibit distinct properties of discourse. Statistically, the size difference between two corpora does not affect the results of keyword extraction (McEnery, Xiao, and Tono 2006, 308-11). The mainstream news corpus, focusing on North Korean women, predominantly includes keywords associated with political relations and incidents related to North Korea. There is a significant presence of nominals with negative semantic prosody, such as 출신 *chulsin* "origin," 가명 *gamyeong* "alias," 공작원 *gongjakwon* "spy," 요원 *yowon* "agent," 단속 *dansok* "crackdown," 제재 *jejae* "sanction," and 납치 *napchi* "kidnap." These keywords collectively shape negative discourse prosody, highlighting themes of espionage, enforcement, and geographical tension.

In contrast, the *Women News* corpus prominently features keywords pertaining to gender issues, particularly focusing on sexual violence and crimes. This corpus includes nominals with negative discourse prosody, including 성폭력

Table 7. Fifty Top Nominal Keyword Lists of Mainstream News and *Women News* (Ranked by Log-Likelihood Scores)

Category	Top fifty nominal keywords
Five mainstream newspapers	탈북자 "NKD," 북한 "NK," 중국 "China," 김정은 "Kim Jong-un," 미국 "US ," 김정일 "Kim Jong-il," 평양 "Pyongyang," 당국 "authorities," 국경 "border," 대사관 "embassy," 대북 "against the North," 소식통 "source," 식당 "restaurant," 수용소 "camp," 탈북민 "NKD," 공안 "Chinese police," 돈 "money," 식량"" 남쪽 "the South," 김일성 "Kim Il-sung," 조선족 "Korean Chinese," 방송 "broadcast," 관계자 "person concerned," 정권 "regime," 망명 "exile," 라오스 "Laos," 김정남 "Kim Jeong-nam," 경찰 "police," 장성택 "Jang Sung-taek," 베이징 "Beijing," 가명 "alias," 남한 "South Korea," 사실 "fact," 당국자 "authorities," 조사 "inspection," 단속 "crackdown," 제재 "sanction," 북쪽 "the North," 출신 "origin," 공작원 "spy," 일본 "Japan," 납치 "kidnap," 브로커 "broker," 정보 "information," 탈출 "escape," 요원 "agent," 총영사관 "general consulate," 현지 "local," 친척 "relative," 대남 "against the South"
Women News	여성 "female," 평화 "peace," 평등 "equality," 성 "sex/gender," 양성 "both sexes," 성폭력 "sexual violence," 이탈 "escape," 젠더 "gender," 정책 "policy," 삶 "life," 여성운동 "women's movement," 학습 "learning," 가족부 "ministry of (women and) family," 성평등 "gender equality," 청소년 "youth," 단절 "disconnection," 포럼 "forum," 사회 "society," 총재 "president," 예방 "prevention," 서울시 "Seoul," 지원 "support," 양립 "compatibility," 문화 "culture," 협의회 "association," 강화 "reinforcement," 가정폭력 "domestic violence," 성희롱 "sexual harassment," 향상 "improvement," 의원 "law maker," 센터 "center," 사회참여 "civic involvement," 여성학 "department of women's studies," 참여 "participation," 단체 "organization," 차별 "discrimination," 폭력 "violence," 한국여성정책연구원 "Korean Women's Development Institute," 회원 "member," 상담 "consulting," 관점 "perspective," 가정 "home," 과정 "process," 경력 "career," 성범죄 "sexual crime," 야시장 "night market," 국회 "congress," 국가행동계획 "national action plan," 저출생 "low birth"

Source: Authors

seongpoklyeok "sexual violence," 성희롱 *seonghuilong* "sexual harassment," 성범죄 *seongbeomjoe* "sexual crime," 가정폭력 *gajeongpoklyeok* "domestic violence," 차별 *chabyeol* "discrimination," and 폭력 *poklyeok* "violence." These keywords indicate the vulnerability of NKD women, emphasizing the dangers they face in various contexts, including their lives in North Korea, their escapes through China, and their subsequent resettlement in South Korea.

Another set of keywords, however, implicates positive semantic prosody on promoting women's civic engagement and dynamic involvement through various organization and groups (e.g. 사회참여 *sahoe chamyeo* "civic engagement," 참여 *chamyeo* "participation," 지원 *jiwon* "support," 향상 *hyangsang* "improvement," 집단 *jipdan* "group," 단체 *danche* "organization," 협의회 *hyeobuihoe* "association," 한국여성정책연구원 *hanguk yeoseong jeongchaek yeonguwon* "Korean Women's Development Institute," and 가족부 *gajokbu* "ministry of women and family"), which align with keywords of feminism and policy (양성 *yangseong* "gender,"

Table 8. Top Thirty Noun Collocates of 탈북 여성 *talbuk yeoseong* "Female Defector"

Category	Top thirty noun collocates of 탈북 여성 "female defector"
five mainstream newspapers	인신매매 "human trafficking," 남성 "male," 중국인 "Chinese," 아이 "kid," 박사 "PhD," 가명 "alias," 자녀 "children," 인권 "human rights," 강제 "enforcement," 결혼 "marriage," 실태 "reality," 삶 "life," 아들 "son," 매매 "trading," 성 "sex/gender," 탈북자 "NKD," 성폭력 "sexual violence," 농촌 "countryside," 남편 "husband," 공안 "Chinese police," 브로커 "broker," 피해 "damage," 방송 "broadcasting," 이야기 "story," 배우자 "spouse," 경찰 "police," 업체 "agency," 문제 "problem," 고통 "pain," 음란 "vulgarities"
Women News	인권 "human rights," 지원 "support," 박사 "PhD," 성폭력 "sexual violence," 피해 "damage," 가정폭력 "domestic violence," 문제 "issue/problem," 소외 "isolation," 영화 "movie," 인신매매 "human trafficking," 정착 "settlement," 상담 "consult," 취업 "employment," 관심 "interest," 일자리 "job," 이주 "migration," 노인 "elders," 안전 "safety," 가정 "family," 성 "sex," 센터 "center," 부모 "parents," 사회 "society," 교육 "education," 사람 "person," 포럼 "forum," 단체 "organization," 삶 "life," 탈북자 "NKD"

Source: Authors

평등 *pyeongdeung* "equality," 성평등 *seongpyeongdeung* "gender equality," 여성운동 *yeoseong undong* "women's movement," 젠더 *jendeo* "gender," 정책 *jeongchaek* "policy," and 국가행동계획 *gukgahaengdonggyehoek* "national action plan") and career issues (경력 *gyeonglyeok* "career" and 단절 *danjeol* "discontinuity"). These keywords signify that *Women News* delivers more active discussions and conveys deeper concern about NKD women's daily lives in South Korea.

Beyond comparisons of keywords, collocations play a vital role in determining discourse properties and tones through the syntagmatic relationships between words. A collocate is defined as a word that frequently appears within a specific proximity to another word. The strength of these associations can be quantified through statistical analyses of corpus data, providing a measure of the likelihood that the co-occurrence of two words is not coincidental (Baker, Hardie, and McEnery 2006, 36-38). This approach is particularly valuable in discourse analysis, as it can uncover the ideological dimensions of language use.

From the mainstream news corpus and the *Women News* corpus, we identified the collocates of 탈북 여성 *talbuk yeoseong* "female defector" within a five-word span using log-likelihood scores. We present the top thirty nominal collocates that co-occur with "female defector" in more than five different articles, excluding proper nouns, pronouns, and numerals, in Table 8. In our corpus data, the word 탈북 여성 *talbuk yeoseong* "female defector" occurs 1,251 times in the NKD women corpus and 391 times in the *Women News* corpus. In contrast to 탈북 여성 *talbuk yeoseong* "female defector," 탈북 남성 *talbuk namseong* "male defector" appears only twenty times in the former and eight times in the latter, posing a challenge for the collocation analysis due to its lower frequency.

Table 8 illustrates that the mainstream media exhibits a more pronounced

negative discourse compared to *Women News*, as evidenced by collocates associated with 인신매매 *insinmaemae* "human trafficking," 가명 *gamyeong* "aliases," 강제 *gangje* "enforcement," 실태 *siltae* "reality," 매매 *maemae* "trading," 성 *seong* "sex," 성폭력 *seongpoklyeok* "sexual violence," 공안 *gongan* "Chinese police," 브로커 *beurokeo* "brokers," 피해 *pihae* "damage," 경찰 *gyeongchal* "police," 업체 *epche* "agencies," 문제 *munje* "problems," 고통 *gotong* "pain," and 음란 *eumlan* "vulgarities." The other collocates fall into categories such as (1) human/family relationships (including terms like 남성 *namseong* "male," 아이 *ai* "child," 결혼 *gyeolhon* "marriage," 삶 *salm* "life," 아들 *ateul* "son," 남편 *nampyeon* "husband," and 탈북자 *talbukja* "NKDs"); (2) physical contexts (농촌 *nongchon* "countryside"); (3) media (방송 *bangsong* "broadcasting" and 이야기 *iyagi* "story"); (4) employment (업체 *epche* "agency"); (5) status (박사 *baksa* "PhD"); and 6) issues (인권 *ingwon* "human rights"). These collocates underscore the narrative surrounding "female defector," emphasizing the adversities faced by NKD women during their escapes from North Korea, typically via China. The focus is on the escape process and the difficulties encountered, rather than on their ongoing challenges or daily experiences in South Korea.

A notable collocate with 탈북 여성 *talbuk yeoseong* "female defector" in Table 8 is 박사 *baksa* "PhD," which has a significant rate of co-occurrence. Of the 285 mentions of 박사 *baksa* "PhD" in the mainstream news corpus, it appears alongside "female defector" thirty-five times across twenty-six different news articles, primarily in reference to Dr. Aeran Lee and Dr. Hyekyung Lee, who are known for being among the first female defectors to earn doctoral degrees. In the *Women News* corpus, 박사 *baksa* "PhD" also stands out as the third most significant collocate, based on log-likelihood statistics, where the collocate "PhD" is linked to "female defector" fifteen times, with thirteen of these occurrences distinctly referring to Dr. Aeran Lee. This distributional fact highlights a unique aspect of media coverage, focusing on the academic and professional achievements of female defectors. McConnell-Ginet (2003, 77-78) defines name calling as an utterance of a characterizing expression directed at an addressee, of which function is to paste the evaluative label on the addressee. The most frequent address term for Dr. Aeran Lee, which is 탈북 여성 1호 박사 *talbuk yeoseong 1 ho baksa* "the first NKD woman PhD," is grammatically optional, but socially loaded. This name calling collocate of "female defector" prevailing in the media ironically attests to the low social status of other NKD women in South Korean society.

The *Women News* corpus distinguishes itself with a broader array of collocates, including 지원 *jiwon* "support," 정착 *jeongchak* "settlement," 상담 *sangdam* "consult," 취업 *chwieop* "employment," 관심 *gwansim* "interest," 일자리 *iljari* "job," 안전 *anjeon* "safety," 센터 *senteo* "center," 사회 *sahoe* "society," 교육 *gyoyuk* "education," 사람 *saram* "person," 포럼 *poreom* "forum," 단체 *danche* "organization," 삶 *salm* "life," and 탈북자 *talbukja* "NKD." These collocates

contribute to a positive discourse tone, as illustrated in examples (5) and (6).

> (5) 여성과 아동이 안전한 사회, 장애 여성·탈북 여성 등 취약소외계층 지원 강화 등을 역점 추진 사업으로 꼽았다 "They cited them as a priority, including projects on a safe society for women and children and strengthening *support* for vulnerable marginalized groups such as women with disabilities and North Korean defectors." (*Women News* September 16, 2011)
> (6) 탈북 여성의 장점과 취업 성공 사례 등을 알려 탈북 여성들의 취업 문을 점차 넓혀나갈 계획이다 "We plan to gradually open the door for North Korean women to enter the workforce by sharing their strengths and success stories." (*Women News* April 8, 2011)

The term 탈북 여성 *talbuk yeoseong* "female defector" in the *Women News* corpus is also associated with collocates that underscore challenges and issues, reflecting a negative discourse prosody. The collocation analysis presents collocates of negative discourse prosody, such as 성폭력 *seongpoklyeok* "sexual violence," 피해 *pihae* "damage," 가정폭력 *gajeongpoklyeok* "domestic violence," 문제 *munje* "problem," 소외 *sooe* "isolation," and 인신매매 *insinmaemae* "human trafficking." These collocates suggest that the portrayal of 탈북 여성 *talbuk yeoseong* "female defector" is intertwined with a negative discourse tone, emphasizing themes of oppression and adversity, as demonstrated in examples (7) and (8).

> (7) 탈북 여성의 성폭력 문제가 생기는 배경에 대해 탈북 여성이 하나원에서 처음 나와 만나는 사람이 신변 보호 담당관으로 … 남한 경찰에 대해서도 북한과 같다고 생각한다는 것이다 "The problem of sexual violence against North Korean women is that the first person they meet when they come out of Hanawon is a personal protection officer... so they think the South Korean police are the same as in North Korea." (*Women News* August 7, 2020)
> (8) 대부분은 중국을 거쳐 남한으로 입국하고 있으며, 탈북 과정에서 중국인이나 조선족과의 매매혼을 비롯해 성폭력, 성매매 등 남성들보다 훨씬 많은 위험 상황에 놓인다 "Most NKD women enter South Korea through China, and face many more risks than their NKD men, including sexual violence and sex trafficking, as well as forced marriages to Chinese or ethnic Koreans in China." (*Women News* January 5, 2018)

While these collocates are utilized to represent the traumatic experiences of many NKD women, they could also influence identity formation of female defectors in public discourse. Stubbs (1996, 195) argues that the repeated use of such collocations and fixed phrases can lead to uncritical acceptance of these characterizations. Our concordance search further reveals that these negative collocates often address general gender violence or the experiences of NKD women in North Korea or China. Of the 340 sentence samples containing the keywords 성폭력 *seongpokryok* "sexual violence" and 성폭행 *seongpokhaeng* "sexual assaults" in *Women News*, 308 sentences pertain to general womanhood, while nineteen target NKD women in North Korea or China and only thirteen

samples concern NKD women in South Korea. As examined with the frequent usages of피해자 *pihaeja* "victim" in (4), juxtaposing NKD women with other marginalized groups (such as women with disabilities, victims of sexual violence, and comfort women) positions them as an otherized class in South Korea. In contrast, 탈북 남성 *talbuk namseong* "NKD men" does not appear in similar juxtapositions.

In addition to negative collocates of "female defectors," the reports of sexual crimes targeting NKD women often depict them as agentless subjects in the context of human trafficking or sexual violence. This is marked by employing passive verb forms, such as 당하다 *danghada* "to be subjected to" or 되다 *doeda* "to become." Grammatically, the NKD women more frequently appear in the object position as a receiver of support, help, and protection. Even in the subject position, as in (10), a female defector is not an agent of action but a theme that is affected by action or events. Such linguistic representations in *Women News* position NKD women as vulnerable victims lacking agency and as objects who require support and help, as exemplified in (9) and (10).

> (9) 여성과 아동이 안전한 사회를 구축하고 장애 여성, 성폭력 피해자, 일본군 위안부 피해자, 탈북 여성 등 사회 소외계층 지원도 강화할 것이다 "We will build a safe society for women and children and strengthen support for marginalized groups, including women with disabilities, victims of sexual violence, Japanese military comfort women, and NKD women." (*Women News* September 11, 2011)
>
> (10) 여성과 아동이 안전한 사회, 장애 여성·탈북 여성 등 취약소외계층 지원 강화 등을 역점 추진 사업으로 꼽았다 "The key initiatives include a safe society for women and children and strengthening support for vulnerable groups such as women with disabilities and NKD women." (*Women News* September 16, 2011)

Persistent media portrayal of NKD women lacking agency risks distorting their identities and oversimplifying public perceptions of their resilience and complex experiences. Kim (2022) explores an intriguing dimension of agency in the context of NKD women's memoirs and highlighted the agency of these women beyond the confines of the colonial gaze and commercial needs. By investigating the subaltern stories of North Korean migrant women living and working clandestinely in China, Choi (2014) also identifies the agency of NKD women in negotiating and making choices for survival under life-threatening situations. These scholarly and intellectual insights challenge Western depictions of NKD women as powerless objects of human trafficking. Moreover, these new perspectives indicate a need for new linguistic representations of the NKDs and NKD womanhood in public discourse, acknowledging their strong will and agency in striving for survival and resettlement in the South. This adoption would advocate for a more positive linguistic representation and a more empowered positioning of these women.

Lastly, we note that the first-person plural pronoun 우리 *uli* "we/our/

us" is in the list of collocates for 여성 *yeoseong* "women" but not for 탈북 여성 *talbuk yeoseong* "female defector" in *Women News*. In general, the usage of 우리 *uli* reflects connectedness with collectivism in Korean society, representing a collective identity, as in (11).

> (11) "…탈북 여성을 통해 우리 근현대사 슬픔의 근원이자 씨앗을 짚는 일, 전쟁을 겪은 세대의 작가인 제가 지금 꼭 해야만 하는 일이라 느꼈습니다" "'…it felt like something I needed to do now, as a writer from a generation that lived through war, to get to the root and seed of our modern-day grief through NKD women." (*Women News* October 14, 2011)

In (11), the first-person pronoun 우리*uli* "we" advocates for the inclusion of female defectors into the ingroup, signaling the speaker's solidarity, closeness, care, or commitment towards the addressee or reader. However, the function of 우리*uli* is complex, encompassing both inclusive and exclusive meanings and referring to various groups: the speaker and the addressee, the speaker with others but not the addressee, or a group that includes the speaker, the addressee, and others (Du Bois 2012, 320). For example, the referent of 우리 *uli* excludes the NKD women in (12).

> (12) 그들을 위무하는 일, 그들을 우리와 하나로 받아들이고 정서적으로 일체감을 가지는 일… "To comfort them, to accept them (female defectors) as one of *us*, to have an emotional connection with them…" (*Women News* September 2, 2011)

The meaning of 우리 *uli* shifts to excluding the NKD women in (12). This exclusion is marked by the use of the third-person pronoun 그들 *geudeul* "they." In Korean, psychological distancing is conveyed through third-person pronouns: 그들 *geudeul* signifies an isolating stance, implying an "us versus them" dichotomy, while 이들 *ideul* "these" suggests closer proximity or engagement.

It is evident that both mainstream newspapers and *Women News* seldom use phrases like 우리 탈북 여성 *uli talbuk yeoseong* "our female defectors" or 우리 탈북민 *uli talbukmin* "our NKDs," contrasting with the frequent usage of우리 여성 *uli yeoseong* "our women," 우리 사회 *uli sahoe* "our society," and 우리 국민 *uli gukmin* "our citizens." This absence of collocation implies that NKD women in South Korea are not fully embraced as members of the South Korean community.

Conclusion

This study employs a corpus linguistics-based critical discourse analysis approach to provide an elaborated view of the construction and representation of identity and gender for NKDs and defector women in public discourse. The analysis offers valuable insights into the complex interplay between language, power, media,

and cultural dynamics in public discourse. Using topic indicators, keywords, and collocation with concordance search, we examined the way that female defectors are represented in South Korean mainstream media that is characterized by polarized political stances—conservative and liberal news (*Chosun Ilbo, JoongAang Ilbo*, and *DongA Ilbo,* versus *Hangyoreh* and *Kyunghyang Sinmun*). As a comparison set, we explored linguistic usages in a smaller media outlet with a feminist perspective, *Women News*.

Our findings illuminate the invisibility of NKD women in media representations across mainstream newspapers, regardless of their distinct political alignments. The scrutiny of topic indicators within the NKD corpus unveils prevalent topics such as "problem," "support," "education," and "human rights." However, these indicators conspicuously omit gender-specific topics, despite the gendered nature of NKD migration, which predominantly involves women. This lack of visibility in mainstream media forms a stark contrast to the popular portrayals of young and attractive NKD women on television talk shows, as discussed by Oh (2016) and Lee (2014). This study, therefore, examines the media representations of NKD women through corpus analysis, with an emphasis on the language used in public discourse. Recognizing the mainstream media's limited engagement with gender perspectives, we compared the mainstream news coverage with that of *Women News*, expecting that with its feminist viewpoint *Women News* would more thoroughly address the gender-related challenges faced by female defectors in South Korea.

Our corpus-based analysis confirms that the conservative media often frame NKD women's issues within narratives of national security threats and human rights violations by North Korea. Liberal outlets, however, address a somewhat broader spectrum of issues concerning NKD women and their rights, but the discourse is often intertwined with other minority groups and topics like labor movements, domestic affairs, and political agendas. While these findings align with previous research across various disciplines, our study contributes significant linguistic evidence, both in quantity and quality, regarding media attitudes. It underscores the pivotal role of media language in shaping public discourse and molding collective and individual identities. Our analysis indicates that news outlets, both conservative and liberal, seldom focus on the specific challenges NKD women face in adapting to life in South Korea. Although *Women News* adopts a more positive and inclusive approach towards NKD women from a feminist standpoint, our detailed examination of keywords and collocations reveals a risk of portraying them as passive victims or objects and of potentially leading to further marginalization by associating them with other marginalized groups in South Korea.

Media language is frequently politically charged, leading to representations that can perpetuate stereotypes against NKD women, particularly through the repeated use of keywords and collocates with negative discourse prosody (e.g.

인신매매 *insinmaemae* "human trafficking," 성폭력 *seongpoklyeok* "sexual violence") and otherizing pronouns (그들 *geudeul* "they" versus 우리 *uli* "we"). The public discourse surrounding NKDs and NDK women is inevitably political. However, media discourse that narrowly focuses on defections, while neglecting the real-life challenges of female defectors as legitimate members of South Korean society, can be detrimental. This concern is amplified by the potential misuse of language in the media, which risks promoting an incorrect and biased representation of NKD women in public discourse, thus misleading the public and distorting their perceptions of these women's identities. Our findings highlight the challenges in how North Korean women are depicted in South Korean society and underscores the need for more gender-aware and inclusive language in public discourse.

Acknowledgements

Notes

1. An anonymous reviewer aptly highlighted that news institutions often adopt varying stances depending on the issues they report. This observation aligns with existing scholarship that acknowledges the pronounced partisan slant within Korean journalism, as evidenced by the works of Han (2018), Lee (2008), and Kim (2011), among others. Given this context and the well-documented ideological divides, our study adopts the conventional binary classification of media outlets into "conservative" (*Chosun Ilbo*, *JoongAng Ilbo,* and *DongA Ilbo*) and "liberal" (*Hangyoreh* and *Kyunghyang Sinmun*).
2. The English translation has been added by the authors.
3. An anonymous reviewer noted potential variability in article volume and language use across extended periods within mainstream media (1992-2022) and *Women News* (2009-2022). This variability could be attributed to fluctuations in NKD inflows, alterations in domestic laws, systems, and policies, and significant NKD-related events. Given these complexities and the scope of this study, a comprehensive diachronic analysis of the NKD corpus will be deferred to future research.
4. Topic indicators of the NKD women corpus are very similar to those of the NKD corpus in Table 1. The only difference is that topic indicators on community/individuals outnumber those on issues/activities/circumstances.
5. A token refers to an instance of a word in a text, essentially each word occurrence. A type, on the other hand, is a unique word in the text, disregarding its frequency (Brezina 2018, 39).
6. The process of morpheme analysis and POS-tagging combines lemmatization. Lemmatization condenses various word forms into a single base form, or lemma. In Korean's agglutinative framework, it combines a noun with its particles and a verb stem

with its endings into one lemma. For instance, 주민 *jumin*, "resident" forms 주민-이 *jumin-i* "resident-subject particle," 주민-을 *jumin-ul*, "resident-object particle," and 주민-에게 *jumin-egye* "resident-to," while 가다 *gada* "to go" yields forms like 가-고 ga-go "to go-ENDING," 가-는 ga-neun "to go-PRESENT," and 가면서 ga-myeonseo "to go-while."

References

Andrews, Edga. 1990. *Markedness Theory: The Union of Asymmetry and Semiosis in Language*. London: Duke University Press.

Anthony, Laurence. 2023. *AntConc*. Version 4.2. Tokyo: Waseda University.

Baker, Paul. 2004. "Querying Keywords: Questions of Difference, Frequency, and Sense in Keywords Analysis." *Journal of English Linguistics* 32 (4): 346-59.

Baker, Paul. 2006. *Using Corpora for Discourse Analysis*. London: Continuum.

Baker, Paul. 2008. *Sexed Text: Language, Gender, and Sexuality*. London: Equinox.

Baker, Paul, Andrew Hardie, and Tony McEnery. 2006. *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Baker, Paul, Costas Gabrielatos, Majid KhosraviNik, Michal Krzyżanowski, and Ruth Wodak. 2008. "A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press." *Discourse & Society* 19 (3): 273-306.

Baker, Paul, Costas Gabrielatos, and Tony McEnery. 2013. *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge, UK: Cambridge Press.

Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge, UK: Cambridge University Press.

Bucholtz, Mary, and Kira Hall. 2005. "Identity and Interaction: A Sociocultural Linguistic Approach." *Discourse Studies* 7 (4-5): 585-614.

Cheng, Winnie. 2013. "Corpus-Based Linguistic Approaches to Critical Discourse Analysis." In *The Encyclopedia of Applied Linguistics*, ed. Carol A. Chapelle, 1353-60. Malden, MA: Blackwell.

Choi, Eunyoung. 2010. "Gender, Justice and Geopolitics of Undocumented North Korean Migration." PhD diss., Syracuse University.

Choi, Eunyoung. 2014. "North Korean Women's Narratives of Migration: Challenging Hegemonic Discourses of Trafficking and Geopolitics." *Annals of the Association of American Geographers* 104 (2): 271-9.

Choi, Yu-sook. 2016. "Sinmungisae natanan bukhanitaljumin jichingeo bunseok – jichingeowa gwallyeoneoui gonggieoreul jungsimeuro" [Reference Terms for North Korean Defectors Seen in Newspaper Articles]. *The Journal of Language & Literature* 67: 33-66.

Chung, Byung-Ho. 2008. "Between Defector and Migrant: Identities and Strategies of North Koreans in South Korea." *Korean Studies* 32 (1): 1-27.

Chung, Byung-Ho. 2014. "North Korean Refugees as Penetrant Transnational Migrants." *Urban Anthropology and Studies of Cultural Systems and World Economic Development* 43 (4): 329-61.

Du Bois, Inke. 2012. "Grammatical, Pragmatic and Sociolinguistic Aspects of the First

Person Plural Pronoun." In *Subjectivity in Language and Discourse*, eds. Nicole Baumgarten, Inke Du Bois and Juliane House, 319-38. Bingley, UK: Emerald Group Publishing.

Fairclough, Norman. 1989. *Language and Power*. London: Longman.

Hall, Stuart. 1997. "The Work of Representation." In *Representation: Cultural Representations and Signifying Practices*, ed. Stuart Hall, 15-64. London: Sage.

Han, Jiyoung. 2018. "Partisan Media and Polarized Opinion in South Korea: A Review." In *Digital Korea: Digital Technology and the Change of Social Life*, eds. Wooyeol Shin, Kyung-Hee Kim and Changwook Kim, 77-101. Seoul: Hanul Academy.

Hardt-Mautner, Gerlinde. 1995. *"Only Connect": Critical Discourse Analysis and Corpus Linguistics*. January, University of Lancaster. http://ucrel.lancaster.ac.uk/papers/techpaper/vol6.pdf (accessed April 16, 2024).

Kim, Sung Kyung. 2012. "'Defector,' 'Refugee,' or 'Migrant'? North Korean Settlers in South Korea's Changing Social Discourse." *North Korean Review* 8 (2): 94-110.

Kim, Sung Kyung. 2022. "Geullobeol daejungmunhwawa saengjonjaui iyagi: haeoechulpan talbuk yeoseong sugiui sikminjeok siseongwa jendeohwadoen seosa" [Global Popular Culture and Stories of Survivors: Colonial Gaze and Gendered Narrative in Memoirs of North Korean Women]. *Journal of Korean Women's Studies* 38 (2): 1-32.

Kim, Young Uk. 2011. "Hanguk eollonui jeongpaseonggwa sahoejeok sotongui wigi" [Partisanship of Korean Press and the Crisis of Social Communication]. *Hangukeollonhakhoe simpojium mich semina* [*Symposium and Seminar of Korean Society for Journalism and Communication Studies*] 107-36.

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.

Kwon, Keum Sang. 2013. "Daejungmaechega saengsanhaneun 'ijuyeoseong' jaehyeonui sahoejeog uimi: gyeolhonijumingwa bukanitaljumin TV peurogeuraemeul jungsimeuro" [Media to Produce: Focusing on Marriage Migrant and North Korean Defector in the TV Programs]. *The Journal of Multicultural Society* 6 (2): 39-81.

Lakoff, Robin. 1973. "Language and Woman's Place." *Language in Society* 2 (1): 45-79.

Lee, Jae-Kyeong. 2008. "Hangukui jeoneollijeumgwa sahoegaldeung: galdeungyubalhyeong jeoneollijeumeul geukbokaryeomyeon" [Korean Journalism and Social Conflict]. *Communication Theories* 4(2): 48-72.

Lee, Soo Jung. 2020. "'talbukja'eseo 'sahoejeok gajang'euro: 'jendeohwadoen chogukjeok iju'ui gwanjeomeseo salpyeo bon yeongguk geoju bukanijunamseongdeurui iju gyeongheomgwa sahoe hwaldongui uimi" [From 'North Korea Defectors' to Social Patriarchs': The Meanings of Migration Experience and Social Activities of North Korean Migrant Men in the United Kingdom]. *Contemporary Society and Multiculturalism* 10 (2): 177-213.

Lee, Sun-Min. 2014. "Talbuk yeoseongeun eotteoke malhal su inneunga?: tellebijeon tokeusyo <Ije mannaleo gapnida>(Chaeneol A)e daehan bipanjeok bunseogeul jungsimeuro" [How Can North Korean Women Defectors Speak: A Critical Analysis of Television Talk Show, Now, Going to Meet (Channel A)]. *Media, Gender & Culture* 29 (2): 75-115.

Lim, Jeongsub, and Myungjun Kim. 2015. "Geungeoironbangbeobui gochalgwa eollonhag yeongue juneun hamui: bosusinmungwa jinbosinmunui talbukja gihoekgisa bunseogeul jungsimeuro" [Review of Grounded Theory Method and Its Implications

for Journalism Studies: An Analysis of Feature Stories about North Korean Defectors in Conservative and Liberal Papers]. *Korean Journal of Journalism and Communication Studies* 59 (2): 427-66.

Louw, Bill. 1993. "Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies." In *Text and Technology: In Honour of John Sinclair*, eds. Mona Baker, Gill Francis and Elena Tognini-Bonelli, 157-76. Amsterdam: John Benjamins Publishing.

McConnell-Ginet, Sally. 2003. "'What's in a Name?' Social Labeling and Gender Practices." In *The Handbook of Language and Gender*, eds. Janet Holmes and Miriam Meyerhoff, 69-97. Oxford: Blackwell.

McEnery, Tony, Richard Xiao, and Yuko Tono. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.

Ministry of Unification. 2024. "Recent Status." Ministry of Unification News. https://www.unikorea.go.kr/unikorea/business/NKDefectorsPolicy/status/lately/ (accessed March 20, 2024).

Oh, Won Hwan. 2016. "Talbukja jeongcheseongui jeongchihak: jongpyeoneseoui 'talbukminyeo'ui deungjanggwa 'talbukja' jeongcheseongui byeonhwareul jungsimeuro" [The Politics of Identity of North Korean Refugees]. *Korean Journal of Broadcasting and Telecommunication Studies* 30 (3): 5-41.

Park, Seong-woo. 2007. "Saeteomineuro bureuji maseyo" [Don't Call Them New Settlers]. April 16, Radio Free Asia. https://www.rfa.org/korean/in_focus/nk_defector_term_new_settlers-20070416.html (accessed December 9, 2023).

Phillips, Martin. 1989. *Lexical Structure of Text*. Discourse Analysis Monograph 12. English Language Research, University of Birmingham.

Stubbs, Michael. 1996. *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell.

Stubbs, Michael. 1997. "Whorf's Children: Critical Comments on Critical Discourse Analysis (CDA)." *British Studies in Applied Linguistics* 12: 100-116.

Stubbs, Michael. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. New York: Blackwell.

Tae, Ji-Ho, and In-Sung Whang. 2012. "Tellebijeon tokeusyo <Ije mannareo kapnida> (chaeneol A)ui talbuk yeoseongdeurui sajeok gieok jaeguseong bangsikgwa geu uimie daehayeo" [How Does Television Talk Show, Now, I'm Coming to Meet You (Channel A), Reconstruct North Korean Women Defectors' Personal Memories?] *Korean Journal of Communication & Information* 60: 104-24.

Van Dijk, Teun A. 1988. *News as Discourse*. Hillsdale, NJ: Erlbaum.

Wodak, Ruth. 2012. "Language, Power and Identity." *Language Teaching* 45 (2): 215-33.

Dr. Sun-Hee Lee, Professor of Korean at Wellesley College's East Asian Languages and Cultures department, specializes in corpus linguistics, learner corpora, and discourse analysis. She has authored numerous publications on Korean grammar, corpus analysis, and learner language. Her research focuses on corpus-based analysis of media, gender, and personal narratives, alongside learner corpus research. Email: slee6@wellesley.edu

Dr. Beomil Kang, a Research Professor at Yonsei University's Institute for the Study of Korean

Modernity, specializes in corpus linguistics, language processing, and statistical analysis. He completed his PhD in Language and Information Studies at Yonsei University in which his dissertation, "A Quantitative Study of Language Change in the Korean Newspaper," analyzes *Chosun Ilbo* articles from 1920 to 2019. Email: kangbeomil@gmail.com