English Teaching, Vol. 79, No. 2, Summer 2024, pp. 89-113
DOI: https://doi.org/10.15858/engtea.79.2.202406.89
http://journal.kate.or.kr

A Corpus-Driven Exploration of English Textbook Evolution in North Korea

Younghee Cheri Lee* and Tae-Young Kim**

Lee, Younghee Cheri, & Kim, Tae-Young. (2024). A corpus-driven exploration of English textbook evolution in North Korea. *English Teaching, 79*(2), 89-113.

This study explores the development of English textbooks in North Korea through corpus-based analysis aimed at illuminating the differences between materials produced during the Kim Jong-il and Kim Jong-un regimes. In the context of educational reforms and changing political ideology, this study investigates BNC/COCA-based lexical coverage and the key lexical features of North Korean middle school English textbooks, highlighting the complexity, vocabulary, and readability of the learning materials. The findings revealed that the Kim Jong-un regime had implemented reforms to improve English language education, with increased lexical diversity, textual complexity, and vocabulary exposure. Although no significant differences were found between the two regimes regarding the lexical coverage of textbooks, the Kim Jong-un regime's textbooks exhibited improvements in diversity, readability, and complexity. This study contributes to a broader understanding of the interplay between political ideology and English language education in North Korea, offering insights that have implications beyond the North Korean context and encouraging reflection on the nation-driven educational reform.

Key words: North Korea, English textbooks, corpus, lexical features, token coverage

*First author: Younghee Cheri Lee, Assistant Professor, English Language and Literature, School of Global Communication, Mokpo National University

**Corresponding author: Tae-Young Kim, Professor, Department of English Education, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, Korea, 06974; E-mail: tykim@cau.ac.kr

Received 31 March 2024; Reviewed 22 April 2024; Accepted 1 June 2024

1. INTRODUCTION

Long recognized for its strong emphasis on political indoctrination and adherence to the ruling regime's principles (Lankov, 2015), the education system in the Democratic People's Republic of Korea (DPRK or North Korea) has attracted growing interest with regard to understanding the impact of its political leadership on educational materials, especially in English language education (Ahn & Lee, 2022). This influence has been particularly evident in the textbooks under the regimes of Kim Jong-il and Kim Jong-un (Kim, 2021). The Kim Jong-un regime has updated its English textbooks to align with global standards, significantly altering the learning materials. Recent North Korean news has highlighted practical English education, English debate competitions, and English and science education for gifted students (see Lee & Kim, 2022).

The existing research on this topic contains valuable insights into the characteristics of English textbooks in various contexts, including North Korea and South Korea, and the influence of political ideology on language learning materials. However, several gaps in the literature warrant further investigation. These gaps include a limited comparative analysis of textbook content and approach across the regimes of Kim Jong-il and Kim Jong-un, as well as the need for a more in-depth, corpus-based analysis of North Korean English textbooks that considers a broader range of lexical features. Although some studies have explored the overall content and ideological perspectives presented in these textbooks (Ahn & Lee, 2022; Choi, 2018; Song, 2002), few have investigated the lexical features that shape North Korean students' English-learning experience and influence their English proficiency.

To address these gaps and provide a comprehensive, corpus-based analysis of English textbook reforms in North Korea, this study compares the lexical feature differences between the textbooks issued during the Kim Jong-il and Kim Jong-un regimes. Moreover, most comparative textbook analyses of these two regimes have been published in Korean, limiting the accessibility and dissemination of the research findings beyond the local Korean context; consequently, international scholars and practitioners cannot build on these insights and engage in a broader dialogue on the subject. In order to address this limitation, the current study has employed a robust and versatile corpus-based approach, utilizing tools and resources—such as British National Corpus/ Corpus of Contemporary American English (BNC/COCA) baseword lists—designed explicitly for analyzing lexical coverage and vocabulary levels to assess the English textbooks used under the Kim Jong-il and Kim Jong-un regimes. By examining the lexical features of English textbooks from different regimes, this study aims to provide insights into each regime's priorities and objectives regarding language education.

Thus, this study attempts to critically evaluate North Korea's English education with a specific focus on lexical features by analyzing the English textbooks used in the country's

(junior) middle schools.[1] To address these objectives, the following questions are explored:

1) How do English textbooks from the Kim Jong-il and Kim Jong-un regimes in North Korea differ in lexical coverage based on BNC/COCA core baseword lists?
2) What are the differences in lexical features between English textbooks published during the Kim Jong-il and Kim Jong-un regimes, and what insights can be gained regarding their diversity, readability, and complexity?

This study will not only analyze the evolution of English textbooks in North Korea but will also consider the implications of these findings for educators and policymakers outside of North Korea.

2. LITERATURE REVIEW

2.1. English Language Education in North Korea

The North Korean education system has long been typified by its emphasis on political indoctrination and strict adherence to the ruling regime's principles (Kirkpatrick & Liddicoat, 2019). Consequently, English language education has disseminated the regime's ideology and fostered loyalty to the state (Lankov, 2014, 2015). Research on North Korean English language education has primarily focused on the textbooks' content and ideological perspectives (Kim, 2021; Lee, 2014; Lee & Kim, 2022). These studies have revealed that these textbooks often reflect the political climate and nationalistic goals of the regime in power, with promoting the state's ideology prioritized over fostering genuine language proficiency (Cho, 2014). Furthermore, Kang (2020) has highlighted the strong influence of political leadership on the design and content of English language learning materials, particularly during the regimes of Kim Jong-il and Kim Jong-un.

North Korea introduced the 2013 New Educational Guidelines under Kim Jong-un's regime, revealing the new direction of education policy. These guidelines, serving as a national curriculum, outline the objectives and methods of school education. The regime's education reform efforts were concretized by the 2013 revision of these guidelines. In 2013, the Overall 12-Year Compulsory Education Guidelines were released, and their full implementation was achieved by 2017 (KIUE, 2021). Despite the COVID-19 pandemic, the regime continues its education reform efforts, shifting toward transitioning to a full-scale implementation of the overall 12-year compulsory education system (KINU, 2021).

In terms of school subjects in North Korea, the Kim Jong-un regime's attitude toward

[1] Here, "middle school" refers to the six-year secondary school system in North Korea, with "junior middle school" referring to the first three years and "senior middle school" referring to the remaining three years.

English has dramatically shifted and become increasingly friendly, which starkly contrasts with the Kim Jong-il regime's hostility toward it. The Kim Jong-un regime has updated its English textbooks to meet global standards, significantly departing from the materials used during the Kim Jong-il regime (Lee & Kim, 2022). This change has been attributed to the Kim Jong-un regime's recognition of the importance of English as a global language, with most science and technical texts having English as their medium, and the desire to enhance the country's international presence and status and to establish its sovereignty as normal (Cho, 2014; Yoo & Kim, 2018). Moreover, Kang (2020) noted that the evolving political landscape in North Korea has led to modifications in language education priorities and objectives.

2.2. Lexical Features of English Textbooks in North Korea

Lexical attributes are vital in shaping the learning experience and can affect learners' language competence (Nation, 2013). Factors such as lexical coverage, lexical variety, average word length, and average sentence length are indispensable for assessing the quality and efficacy of language learning resources (Lu, 2012; McCarthy & Jarvis, 2010). Lexical coverage, as gauged by BNC/COCA baseword lists, is a significant parameter that represents the breadth of vocabulary encountered by learners, and it can influence their comprehensive proficiency (Nation, 2006).

Similarly, lexical variety, as measured by the standardized type/token ratio (STTR), has been associated with the richness and intricacy of language materials, which can affect learners' capacity to assimilate new vocabulary and refine their language abilities (Malvern, Richards, Chipere, & Durán, 2004). Average sentence length and average word length also function as crucial indicators of learning materials' readability and complexity, potentially impacting learners' understanding and retention of novel language structures and vocabulary (Lu, 2012; McCarthy & Jarvis, 2010).

For vocabulary profiling, several studies analyzed the vocabulary items in the English textbooks used in North Korean middle schools and compared them to South Korean textbooks using baseword lists. These studies found that, although both sets of textbooks included a considerable amount of high-frequency vocabulary, the North Korean textbooks placed a greater emphasis on political and ideological contents, resulting in a distinct vocabulary profile (see Hwang & Kim, 2017; Kim, Lee, & Kim, 2017). Similarly, recent studies have started to utilize baseword lists, such as the General Service List (GSL) and Academic Word List (AWL), to analyze and compare English textbooks from North Korea and South Korea.[2] These studies have provided valuable insights into the differences and

[2] The General Service List (GSL) comprises approximately 2,000 essential English words chosen for

similarities in the vocabulary content and coverage of the textbooks from these two countries. For instance, Kim (2020) conducted a comparative analysis of North Korean and South Korean English textbooks, employing GSL and AWL with reference to BNC/COCA baseword lists to examine the frequency and distribution of vocabulary items. The study aimed to understand the quantitative and qualitative differences between the textbooks and, ultimately, lay the foundation for creating a unified English vocabulary list for future textbooks. Both North Korean and South Korean textbooks were found to contain a similar number of words equivalent to BNC/COCA core lists; however, North Korean textbooks had more specialized academic vocabulary. The findings revealed notable differences in the coverage of high-frequency words and the inclusion of topic-specific vocabulary; these were attributed to the countries' respective educational goals and sociopolitical contexts.

Kim Jong-un's strong emphasis on education, particularly the English language, as the foundation of science and technology has accelerated these transformations systematically and pragmatically (as cited in Lee & Kim, 2022). A growing body of research is aiming to outline the textual and formatting disparities between English textbooks produced during the regime of Kim Jong-un and those of his predecessors. However, few studies have elucidated how English language input has been employed to reflect educational reform. Lee and Kim (2022) argued that the revised textbooks from Kim Jong-un's regime meet the lexical threshold of the BNC/COCA 3K core vocabulary, which indicates meticulously calibrated language input in terms of both quality and quantity. Furthermore, they found that these revised editions contain a greater proportion of authentic and high-frequency 3-gram lexical bundles than the textbooks from Kim Jong-il's regime. This shift in language input is significant as it demonstrates the Kim Jong-un regime's commitment to enhancing the quality of English language education while maintaining the regime's self-reliance and self-sustainability.

3. METHODS

3.1. Corpus Data

The present study aims to examine the differences in English textbooks between the Kim Jong-il and Kim Jong-un regimes in North Korea with regard to linguistic aspects. A corpus-based quantitative analysis was used to investigate significant shifts in North Korean English education, by examining English textbooks. A representative sample of English textbooks

their widespread use. The Academic Word List (AWL) contains words frequently found in academic texts across various subjects; it was developed by Averil Coxhead at Victoria University of Wellington, New Zealand, and is useful for English language students.

from both regimes was collected to construct a comparable corpus, including three grade levels from before and after Kim Jong-un came to power, corresponding to middle school textbooks in South Korea. In this study, English textbooks were divided into essential reading texts, supplementary reading texts, directions, and tasks; subsequently, sub-corpora were independently constructed for each grade level under the Kim Jong-il and Kim Jong-un regimes. The study was designed as the Comparable Corpora of English in North Korea (CCENK) 2.0, and only the essential reading sections of the sub-corpora were included and analyzed.[3]

The selected grade levels and textbooks corresponded to South Korean middle school Grades 1 to 3 and their respective textbooks. Grades 1 to 3 textbooks from the late Kim Jong-il and Kim Jong-un regimes were chosen as the corpus construction data. The sampled textbooks were digitized, converted into a machine-readable format, and yielded two separate corpora: one for the Kim Jong-il regime and the other for the Kim Jong-un regime, which enabled the corpus-based analysis of the texts. Each corpus contained only the textual content from the selected textbooks, excluding images, tables, and other non-textual elements. Subsequently, the corpora were preprocessed, which involved the removal of special characters and other irrelevant symbols. The overall scale of the constructed North Korean English textbook CCENK corpus is presented in Table 1.

TABLE 1

The Size of CCENK 2.0

Group	Sub-Corpus	Equivalent to South K. Grade	Unit (#)	Splitted (#)	Token (#)	Type (#)	Sentence (#)	MSL (#)
Kim Jong-il *Before Revision*	JI_M1	Middle School Grade 1	19	30	939	286	219	4.29
	JI_M2	Middle School Grade 2	16	30	927	348	182	5.09
	JI_M3	Middle School Grade 3	17	30	2,060	496	281	7.33
Kim Jong-un *After Revision*	JU_M1	Middle School Grade 1	15	30	525	238	114	4.61
	JU_M2	Middle School Grade 2	12	30	1,024	452	159	6.44
	JU_M3	Middle School Grade 3	12	30	1,398	543	181	7.72

**Notes:* In both regimes, the sub-corpus name was created using the initials of their first names, JI for Kim Jong-il and JU for Kim Jong-un.

It should be noted that the length of reading passages and the number of sentences in

[3] CCENK 1.0, initially developed for a project comparing high school English textbooks between the Kim Jong-il and Kim Jong-un regimes, is detailed in Lee and Kim (2022).

English textbooks from Kim Jong-il's regime exhibited a considerable difference compared to those from Kim Jong-un's regime. Furthermore, these numerical values suggested that the amount of learning content was not evenly distributed across grade levels, considering the volume of learning material, during Kim Jong-il's regime. In contrast, during Kim Jong-un's regime, the discrepancies in these numerical indices were apparently minimized across grade levels. The significance of these differences will be further elaborated in the Results section.

3.2. Encoded Parameters

To investigate the lexical features of the English textbooks, the following parameters were employed: a comparison based on BNC/COCA 1K to 34K baseword lists for lexical coverage; STTR for lexical diversity; and mean word length (MWL) and mean sentence length (MSL) for textual complexity. The analysis of lexical coverage was conducted using AntWordProfiler 3.5.9 (Anthony, 2020), a software tool designed for profiling the vocabulary levels of texts; it was further customized based on BNC/COCA baseword lists.

Lexical Coverage: BNC/COCA Baseword Lists. As the initial step in the analyses, the difference in lexical coverage ratios was measured by examining the vocabulary profiles of the texts in English textbooks from before and after Kim Jong-un came to power. Generally, by analyzing vocabulary profiles, it is possible to obtain statistical information on the scale and level of the vocabulary comprising a given text. Instead of using the default lists, the AntWordProfiler program was customized with BNC/COCA 1K to 34K vocabulary lists, which suited the purpose of this study.

The BNC/COCA list is a continually updated and modified vocabulary list, first compiled by Nation in 2012 (Nation, 2016, 2017); it combines his 2005 version of the BNC list with Mark Davies' COCA (Davies, 2008), which had been updated until 2012. Here, 1K represents a word family of 1,000 headwords, with lists from 1K to 25K compiled based on frequency and range. Lists from 26K to 30K are currently incomplete, while lists from 31K to 34K consist of special listings, such as proper nouns, abbreviations, compound words, and onomatopoeic words. A comparison with the frequency-based vocabulary list from the modern BNC was deemed necessary as North Korean English textbooks follow British English rules. Additionally, the COCA frequency-based vocabulary list, updated annually with new corpus compositions, was chosen as it was deemed essential for understanding the linguistic traits of North Korean English textbooks from the perspective of modern English.

Furthermore, to better comprehend the impact of the frequency of extra words on lexical coverage, four additional special vocabulary lists (i.e., BNC/COCA 31K~34K) provided separately from Nation's (2016) BNC/COCA 1K to 25K lists were also examined.[4]

[4] As stated by Nation (2016), BNC/COCA baseword lists 31K to 35K encompass additional words

Token and Type. The preliminary analysis examined the statistical differences between the two regimes using tokens and types to measure lexical richness, diversity, and complexity. Tokens and types are linguistic terms often employed in analyzing text corpora to better understand the diversity and frequency of the vocabulary used in a text. Tokens refer to the total number of individual words in a text, representing its total word count. For example, in the sentence "The cat chased the mouse," there are five tokens, as there are five individual words. Types refer to the number of unique words in a text, without considering their frequency. In the aforementioned example, there are only four types: "the," "cat," "chased," and "mouse." A higher ratio of types to tokens indicates a more diverse vocabulary, while a lower ratio suggests a more repetitive or limited vocabulary.

Lexical Diversity: STTR. Following the preliminary analysis, the lexical variety in English textbooks from before and after Kim Jong-un came to power was examined to compare their lexical features and capture the qualitative differences between the two regimes. The STTR values of three textbooks from before Kim Jong-un's regime were compared with those of three textbooks during his regime, utilizing the vocabulary statistics information provided by the WordList feature in WordSmith Tools 7.0 (Scott, 2016). The STTR is an index that measures the type/token ratio by standardizing the text length when the lengths of various texts differ. The STTR is widely used as a lexical variety indicator in corpus linguistics. Generally, the higher the STTR value compared to the benchmark, the more diverse the vocabulary used and the higher the text quality. Additionally, the higher the STTR index, the closer the text is to the characteristics of native English (nativeness; Lee, 2018, 2019, 2021).

Textual Complexity: MWL and MSL. Two analytical indicators, MWL and MSL, were used to compare textual complexity and capture the qualitative differences between the two regimes. MWL is a metric used in linguistics, text analysis, and readability assessments to determine the average length of the words in a given text or language sample (Lu, 2012). It is typically calculated by dividing the total number of characters (excluding spaces and punctuation) by the total number of words. MSL is utilized to assess the complexity and readability of a text (Crossley, Allen, & McNamara, 2011). It is calculated by dividing the total number of words in a text by the total number of sentences.

In L2 teaching, using graded materials with controlled vocabulary and sentence structures can help ensure that the content is accessible to students at a particular proficiency level (Nation, 2009). As students make progress through the levels, the MSL in these materials

---

that deviate from standard-level lists, as they consist of specific, purpose-driven words for each category. For instance, BNC/COCA 31K focuses on proper nouns (e.g., Matilda); BNC/COCA 32K features marginal words such as onomatopoeic and mimetic words (e.g., Gee); BNC/COCA 33K includes compound words (e.g., Airforce); and BNC/COCA 34K contains acronyms (e.g., ATM). These baseword lists are also constructed based on word families (Lee & Kim, 2022).

may increase to expose them to more complex target language structures and ideas. A text's content and context can also influence the MSL in middle school English textbooks and materials. For example, texts related to science or technical subjects may have a longer MSL due to the use of more complex language structures and ideas (Biber, 1991).

Although specific research on MSL in North Korean middle school English textbooks or materials is limited, it is generally understood that readability, language complexity, and graded materials are important factors to consider when evaluating and designing educational materials for L2 learners. In the present study, both these parameters were computed using WordSmith Tools 7.0 (Scott, 2016).

4. RESULTS

The lexical coverage comparison was conducted using a three-tiered approach. In the first tier, BNC/COCA core baseword lists ranging from 1K to 3K were each processed separately. These core words are typically considered essential English vocabulary for L2 learners, regardless of their learning context. The first three core baseword lists were selected to evaluate whether they satisfy the fundamental components of L2 learning and to determine whether North Korean middle school English textbooks meet the universal vocabulary threshold standards, as seen in South Korea. Second, statistical analyses were performed to identify group differences between the regimes using BNC/COCA core lists. Finally, BNC/COCA baseword lists from 1K to 34K were analyzed to observe linguistic behavior across the comprehensive and qualitative aspects of the CCENK corpus.

The lexical feature comparison was conducted in a twofold manner. The first phase served as a preliminary analysis to facilitate an overall understanding of the CCENK corpus. At this stage, the statistical differences between the groups were investigated using tokens and types as encoded factors. In the second phase, the numerical values of STTR, MWL, and MSL were compared between the Kim Jong-il and Kim Jong-un regimes.

4.1. Lexical Coverage Comparison

First, different levels of vocabulary distribution were observed using individual BNC/COCA 1K to 3K baseword lists. The results showed that the BNC/COCA 1K portion in Kim Jong-il Middle School Grade 1 (JI_M1) was higher than that in Kim Jong-un Middle School Grade 1 (JU_M1), implying that Kim Jong-il's Grade 1 textbook may contain easier vocabulary words than Kim Jong-un's textbook.[5]

[5] The acronyms JI_M1 and JU_M1 signify the first grade of a middle school under Kim Jong-il's

F-tests and *t*-tests were conducted to identify group differences further to assess the statistical significance of the observed differences for each grade between the two regimes. The lexical coverage comparison between the two regimes revealed marginally significant differences only in Grade 1, showing only statistical insignificance for the other grades' textbooks (JI_M1 vs. JU_M1: $p = .026$*; JI_M2 vs. JU_M2: $p = .217$; JI_M3 vs. JU_M3: $p = .171$). The results of these tests are presented in Table 2, with their significant differences.

TABLE 2[6]

Lexical Coverage Comparison: BNC/COCA 1K

Grade	Group	*F*-test		*t*-test				
		F	*p*	*N*	*M*	*VAR*	*t*	*p*
Grade 1 (M1)	JI_M1	0.421	.011*	30	85.600	70.751	1.999	.026*
	JU_M1			30	80.172	167.96		
Grade 2 (M2)	JI_M2	1.394	.188	30	84.982	102.361	0.788	.217
	JU_M2			30	83.076	73.411		
Grade 3 (M3)	JI_M3	1.138	.365	30	88.798	29.436	0.956	.171
	JU_M3			30	87.500	25.874		

*$p < .05$, **$p < .01$, and ***$p < .001$

The analysis involved performing the same t-test after an F-test, using BNC/COCA 2K to evaluate the differences between the two groups. The outcome of the t-test revealed no statistically significant disparity, as all the p-values exceeded the 0.05 threshold (JI_M1 vs. JU_M1: $p = .480$; JI_M2 vs. JU_M2: $p = .409$; JI_M3 vs. JU_M3: $p = .446$). Thus, no meaningful difference existed between the means of each set of paired groups (See Table 3).

leadership and Kim Jong-un's leadership, respectively, in North Korea.

[6] In our statistical analysis, detailed raw *p*-values are presented for both *F*-tests and *t*-tests to maintain transparency. The notation "$p < .05$" is used alongside the *F*-test results to facilitate the reader's understanding of the statistical process. This notation indicates that, following the *F*-test, the data met the criteria for homogeneity of variances ($p > .05$), and hence, equal variance *t*-tests were subsequently performed. This stepwise approach in the reporting of our statistical results is designed to aid in the clear interpretation of the analysis without compromising the integrity of the statistical reporting.

TABLE 3

Lexical Coverage Comparison: BNC/COCA 2K

Grade	Group	*F*-test		*t*-test				
		F	*p*	*N*	*M*	*VAR*	*t*	*p*
Grade 1 (M1)	JI_M1	0.586	.078	30	3.081	11.284	-0.053	.480
	JU_M1			30	3.134	19.273		
Grade 2 (M2)	JI_M2	3.950	.000***	30	4.354	70.219	-0.231	.409
	JU_M2			30	4.750	17.776		
Grade 3 (M3)	JI_M3	1.366	.203	30	4.064	13.073	0.137	.446
	JU_M3			30	3.945	99.569		

*p < .05, **p < .01, and ***p < .001

The same analysis was also performed using BNC/COCA 3K to assess the differences between the two groups. A statistically significant disparity was demonstrated, with the p-value being lower than the 0.05 threshold (p = .046) only in Grade 1 (JI_M1 vs. JU_M1: p = .046*; JI_M2 vs. JU_M2: p = .118; JI_M3 vs. JU_M3: p = .104). This finding indicated a significant difference between the means of the two groups. The results imply that Kim Jong-un's Grade 1 textbook may contain more vocabulary words from BNC/COCA 3K than Kim Jong-il's textbook, indicating that the former likely has a greater portion of difficult words than the latter (See Table 4).

TABLE 4

Lexical Coverage Comparison: BNC/COCA 3K

Grade	Group	*F*-test		*t*-test				
		F	*p*	*N*	*M*	*VAR*	*t*	*p*
Grade 1 (M1)	JI_M1	0.403	.009**	30	1.349	6.209	-1.716	.046*
	JU_M1			30	2.806	15.401		
Grade 2 (M2)	JI_M2	1.350	.219	30	3.310	15.510	1.196	.118
	JU_M2			30	2.176	11.487		
Grade 3 (M3)	JI_M3	1.233	.288	30	3.269	9.143	1.275	.104
	JU_M3			30	2.322	7.416		

*p < .05, **p < .01, and ***p < .001

Despite the limitation of not having a sufficient sample size collected from North Korean middle-school English textbooks, the aforementioned results are intriguing; this is particularly the case when juxtaposing these findings with our prior investigation into lexical disparities found in North Korean high school English textbooks between the two regimes, which revelated statistically significant discrepancies between them.

Building on the 95% lexical threshold level, further quantitative analyses by grade, as well as supplemental qualitative investigations, were conducted to enhance the initial analysis performed with individual BNC/COCA baseword lists. The first three core baseword lists,

namely BNC/COCA 1K to 3K, and additional lists, such as BNC/COCA 31K to 34K, were examined to determine if they could satisfy the 95% lexical threshold level. Table 5 compares the lexical coverage of Grade 1 English textbooks from the Kim Jong-il and Kim Jong-un regimes. Notably, neither group achieved a text coverage level of 95%, even when considering BNC/COCA 5K. The results indicate that North Korean middle school English textbooks do not adequately meet the universal vocabulary threshold standards like those in South Korea (see Lee & Kim, 2023).

Nonetheless, a closer qualitative examination revealed that Kim Jong-un's textbooks may be perceived as more accessible than Kim Jong-il's textbooks due to their active use of foreign loan words in Korean. For example, the word "sweater" in North Korean usage is written as "스웨타"; this is a transliteration that closely resembles the English pronunciation, making it more accessible to North Korean learners. This finding suggests that, although the lexical coverage figures for the English textbooks of both regimes may seem somewhat similar, the perceived difficulty may be lower in Kim Jong-un's textbooks. In essence, the lexical coverage comparison indicates that students encounter more diverse vocabulary under the Kim Jong-un regime but that a considerable number of these are foreign loan words. Thus, the actual level of difficulty under the Kim Jong-un regime may be similar to or even perceived as lower than the level of difficulty under the Kim Jong-il regime.

TABLE 5

Lexical Coverage Comparison in Grade 1:

95% Lexical Threshold Level Using BNC/COCA 1K~34K

	BNC/COCA 1K~34K	JI_M1		BNC/COCA 1K~34K	JU_M1	
Level	Baseword List	Token (%)	Cum. Token (%)	Baseword List	Token (%)	Cum. Token (%)
	31K_Proper Noun	3.24	3.24	31K_Proper Noun	5.98	5.98
	32K_Marginal	2.48	5.72	32K_Marginal	2.91	8.89
	33K_Compound	0.57	6.29	33K_Compound	0.51	9.40
	34K_Acronym	0.00	6.29	34K_Acronym	1.88	11.28
1	1K	84.75	91.04	1K	78.36	90.08
2	2K	2.86	93.90	2K	3.07	93.16
3	3K	0.00	93.90	3K	0.34	93.50
4	4K	0.10	94.00	4K	0.51	94.01
5	5K	0.67	94.67	5K	0.85	94.69
8	Non-Level	5.34	100.01	Non-Level	5.30	99.99

Table 6 shows a stark difference in the lexical threshold level between the Kim Jong-il and Kim Jong-un regimes' Grade 2 textbooks. The results showed that the Kim Jong-il regime's textbook satisfied a lexical coverage of 95% at BNC/COCA 5K, whereas the Kim Jong-un regime's textbook nearly reached 95% at BNC/COCA 3K. Notably, despite the

appearances of "accordion" from BNC/COCA 10K and "pater" from 16K in the Kim Jong-il regime's textbook, the analysis showed that the lexical coverage fell significantly below the 98% mark. Furthermore, in the Kim Jong-un regime's textbook, "cinema" from 5K, "tram" from 7K, and "noodle" from 8K all appeared once. Overall, the revised textbooks under Kim Jong-un's regime appear to have been compiled at a more accessible level in terms of vocabulary quantity and variety, suggesting that learners may find the vocabulary in the updated textbooks more accessible and comprehensible. As revealed through the lexical coverage analysis, more vocabulary appears in the textbooks under the Kim Jong-un regime; however, as Table 6 indicates, when considering that the threshold for BNC/COCA 3K is being met, it can be seen that they have tried to present a greater amount of highly frequent English vocabulary, with greater variety.

TABLE 6

Lexical Coverage Comparison in Grade 2:

95% Lexical Threshold Level Using BNC/COCA 1K~34K

	BNC/COCA 1K~34K	JI_M2		BNC/COCA 1K~34K	JU_M2	
Level	Baseword List	Token (%)	Cum. Token (%)	Baseword List	Token (%)	Cum. Token (%)
	31K_Proper Noun	2.37	2.37	31K_Proper Noun	3.03	3.03
	32K_Marginal	2.17	4.54	32K_Marginal	2.09	5.12
	33K_Compound	0.59	5.13	33K_Compound	1.36	6.48
	34K_Acronym	0.00	5.13	34K_Acronym	1.36	7.84
1	1K	82.92	88.05	1K	80.86	88.70
2	2K	3.95	92.00	2K	5.02	93.72
3	3K	1.18	93.18	3K	1.26	94.98
4	4K	1.18	94.36	4K	0.10	95.98
5	5K	0.79	95.15	5K	0.63	95.71
6	10K	0.10	95.25	6K	0.00	95.71
7	16K	0.10	95.35	7K	0.10	95.81
				8K	0.10	95.91
8	Non-Level	4.64	99.99	Non-Level	4.08	99.99

Table 7 presents a striking difference in Grade 3 English textbooks between the two regimes. The Kim Jong-il regime's textbook satisfied a lexical coverage of 95% at BNC/COCA 4K, while the Kim Jong-un regime's textbook reached the 95% level at BNC/COCA 3K. The Kim Jong-il regime's textbook failed to meet the 98% mark even at BNC/COCA 13K.

As the threshold level was met at BNC/COCA core lists comprised of relatively easier vocabulary words in Kim Jong-un's regime compared to Kim Jong-il's regime, overall, the vocabulary level and variety in the revised textbooks under Kim Jong-un's regime are considerably more accessible than those under Kim Jong-il's regime.

However, as Lee and Kim's (2022) previous research found that the vocabulary level for Grade 1 high school students was established at BNC/COCA 3K, the overall difficulty level of middle school English textbooks may be perceived as high. Furthermore, according to Lee and Kim (2023), South Korean Grade 3 middle school English is formed at the BNC/COCA 2K~3K level, suggesting a noticeable discrepancy in English lexical input between North Korea and South Korea.

TABLE 7

Lexical Coverage Comparison in Grade 3:

95% Lexical Threshold Level Using BNC/COCA 1K~34K

	BNC/COCA 1K~34K	JI_M3		BNC/COCA 1K~34K	JU_M3	
Level	Baseword List	Token (%)	Cum. Token (%)	Baseword List	Token (%)	Cum. Token (%)
	31K_Proper Noun	1.98	1.98	31K_Proper Noun	2.77	2.77
	32K_Marginal	1.21	3.19	32K_Marginal	1.35	4.12
	33K_Compound	0.63	3.82	33K_Compound	0.28	4.4
	34K_Acronym	0.00	3.82	34K_Acronym	0.92	5.32
1	1K	86.05	89.87	1K	85.1	90.42
2	2K	4.05	93.92	2K	4.05	94.47
3	3K	0.97	94.89	3K	0.57	95.04
4	4K	0.29	95.18	4K	1.42	96.46
5	5K	0.77	95.95	5K	0.50	96.96
6	6K	0.87	96.82	6K	0.00	96.96
7	7K	0.05	96.87	7K	0.00	96.96
8	13K	0.05	96.92	8K	0.07	97.03
	Non-Level	3.09	100.01	Non-Level	2.98	100.01

4.2. Lexical Feature Comparison

Tokens and Types. The statistical differences between the two regimes were analyzed using tokens and types as an initial analysis. On investigating the token values, statistically significant differences were observed between the two groups across all grades (JI_M1 vs. JU_M1: $p < .001$***; JI_M2 vs. JU_M2: $p < .001$***; JI_M3 vs. JU_M3: $p < .001$***). As shown in Table 8, for Grades 1 and 3, the English textbooks used in the Kim Jong-il regime had a substantially more significant number of tokens than those in the Kim Jong-un regime, with some cases even reaching nearly twice the amount; this suggests that North Korean learners may have experienced a heavier learning burden during the Kim Jong-il regime.

Grade 2 exhibited a slightly different pattern. The Kim Jong-un regime's textbooks had a marginally higher number of tokens compared to the Kim Jong-il regime's textbooks. The gradually increasing number of tokens in the revised textbooks under Kim Jong-un's regime seems to have resulted in English textbooks that are more balanced and well organized.

TABLE 8

Token Comparison

Grade	Group	*F*-test		*t*-test				
		F	*p*	*N*	*M*	*VAR*	*t*	*p*
Grade 1 (M1)	JI_M1	2.090	.026*	30	31.3	28.2	11.703	.001***
	JU_M1			30	17.5	13.5		
Grade 2 (M2)	JI_M2	1.203	.311	30	30.9	21.9	-2.793	.001***
	JU_M2			30	34.1	18.6		
Grade 3 (M3)	JI_M3	0.697	.168	30	68.7	26.2	15.125	.001***
	JU_M3			30	46.6	37.6		

*p < .05, **p < .01, and ***p < .001

Additionally, as shown in Table 9, the vocabulary types also demonstrated statistical significance while exhibiting a trend similar to that of tokens (JI_M1 vs. JU_M1: p < .001***; JI_M2 vs. JU_M2: p < .001***; JI_M3 vs. JU_M3: p < .001***).

TABLE 9

Type Comparison

Grade	Group	*F*-test		*t*-test				
		F	*p*	*N*	*M*	*VAR*	*t*	*p*
Grade 1 (M1)	JI_M1	2.594	.006**	30	22.3	22.9	6.959	.001***
	JU_M1			30	15.1	8.9		
Grade 2 (M2)	JI_M2	0.497	.032*	30	23.7	8.1	-4.182	.001***
	JU_M2			30	27.5	16.3		
Grade 3 (M3)	JI_M3	1.379	.196	30	44.6	38.9	5.753	.001***
	JU_M3			30	36.1	28.2		

*p < .05, **p < .01, and ***p < .001

For Grades 1 and 3, the number of types in Kim Jong-un's textbooks significantly decreased compared to those in Kim Jong-il's textbooks. In contrast, for Grade 2, there was a slight increase. These changes can be interpreted as attempts to enhance the quality of North Korean English education, particularly under Kim Jong-un's regime, by carefully controlling the quantity and quality of the English lexical input provided in the textbooks. This calibration of educational materials can be viewed as part of the Kim Jong-un regime's efforts to foster a more effective and engaging student learning experience.

STTR. The analysis of the STTR values in Table 10, measuring the lexical diversity in the textbooks, revealed differences between the two regimes. For the STTR parameter, statistically significant differences were observed between the two groups. Notably, in Grades 2 and 3, the English textbooks utilized during the Kim Jong-un regime exhibited higher lexical diversity values than those utilized during the Kim Jong-il regime, and these differences were statistically significant (JI_M1 vs. JU_M1: p =.407; JI_M2 vs. JU_M2: p

< .001***; JI_M3 vs. JU_M3: p < .001***). However, both regimes displayed similar levels of lexical diversity for Grade 1 textbooks.

These findings suggest that the English textbooks from the Kim Jong-un regime, characterized by higher lexical diversity values, may be considered an improvement in terms of text quality compared to those from the Kim Jong-il regime. As texts with higher STTR values tend to possess better writing quality (Lee & Kim, 2022), it can be stated that the increased lexical diversity in the Kim Jong-un regime's textbooks could potentially facilitate a more enriching and engaging learning experience by exposing North Korean learners of English to a wider variety of words, thereby enhancing their English language education.

TABLE 10

STTR Comparison

Grade	Group	*F*-test		*t*-test				
		F	*p*	*N*	*M*	*VAR*	*t*	*p*
Grade 1 (M1)	JI_M1	2.176	.020*	30	86.722	56.031	0.236	.407
	JU_M1			30	86.333	25.747		
Grade 2 (M2)	JI_M2	2.544	.007**	30	89.556	22.592	-3.145	.001***
	JU_M2			30	92.777	8.879		
Grade 3 (M3)	JI_M3	2.643	.005**	30	92.444	11.988	-2.425	.001***
	JU_M3			30	94.244	4.535		

*p < .05, **p < .01, and ***p < .001

MWL. The MWL was calculated for the English textbooks from both regimes; this revealed variations in the complexity and structure of the texts. As shown in Table 11, statistically significant differences were observed in the average word length parameter between the two groups across all grade levels, accompanied by a general increase in word length (JI_M1 vs. JU_M1: p < .001***; JI_M2 vs. JU_M2: p = .004**; JI_M3 vs. JU_M3: p = .007**). Notably, the difference between the two regimes exhibited a significantly pronounced statistical distinction across all grade levels.

These results, which are in line with the earlier comparison based on the lexical diversity index identified in the STTR analysis, suggest that the English textbooks from the Kim Jong-un regime possess a relative advantage in terms of the quantity of language input and the qualitative aspects. The increased MWL across all grade levels potentially contributes to a more comprehensive and engaging learning experience for students, exposing them to a wider array of vocabulary and thus fostering a deeper understanding and appreciation of English. This ultimately underscores the superiority of the English textbooks from the Kim Jong-un regime compared to the previous regimes.

TABLE 11

Mean Word Length Comparison

Grade	Group	*F*-test		*t*-test				
		F	*p*	*N*	*M*	*VAR*	*t*	*p*
Grade 1 (M1)	JI_M1	1.176	.333	30	3.460	0.230	-3.685	.001***
	JU_M1			30	3.899	0.195		
Grade 2 (M2)	JI_M2	0.920	.412	30	3.790	0.156	-2.720	.004**
	JU_M2			30	4.073	0.170		
Grade 3 (M3)	JI_M3	0.527	.045*	30	3.869	0.062	-2.513	.007**
	JU_M3			30	4.063	0.117		

*$p < .05$, **$p < .01$, and ***$p < .001$

MSL. The MSL was also calculated for the English textbooks from both regimes. As shown in Table 12, statistical significance was found only for Grade 2 (JI_M1 vs. JU_M1: p = .113; JI_M2 vs. JU_M2: p = .001***; JI_M3 vs. JU_M3: p = .164). It was found that, after revisions, the Kim Jung-un regime's textbooks exhibited a slight increase in sentence length compared to the Kim Jung-il regime's ones.

While no statistically significant differences were detected for the other two grade levels, this result indicates that the sentence length differences have been effectively adjusted across grade levels in the revised English textbooks under the Kim Jong-un regime (JU_M1: 5.016, JU_M2: 6.897, JU_M3: 8.217). The analysis suggests that these revised English textbooks have intentionally lengthened the sentences and adequately addressed the issue of sentence-length disparity across different grade levels. These findings demonstrate the efforts made to improve the quality and effectiveness of English education in North Korea under the Kim Jong-un regime.

TABLE 12

Mean Sentence Length Comparison

Grade	Group	*F*-test		*t*-test				
		F	*p*	*N*	*M*	*VAR*	*t*	*p*
Grade 1 (M1)	JI_M1	0.678	.150	30	4.516	2.018	-1.225	.113
	JU_M1			30	5.016	2.977		
Grade 2 (M2)	JI_M2	0.142	.000***	30	5.203	1.109	-3.107	.001***
	JU_M2			30	6.897	7.809		
Grade 3 (M3)	JI_M3	0.483	.027*	30	7.627	3.478	-0.989	.164
	JU_M3			30	8.217	7.207		

*$p < .05$, **$p < .01$, and ***$p < .001$

5. DISCUSSION

5.1. Vocabulary Levels: A Comparison of Lexical Coverage

The analysis presented in this study offers valuable insights into the lexical coverage comparison between North Korean middle school English textbooks published under Kim Jong-il's and Kim Jong-un's regimes. In particular, two statistically significant findings are notable; both were observed only in Grade 1 textbooks. The first one concerned the difference observed in the vocabulary distribution, suggesting that the Kim Jong-il regime's textbooks may contain easier "core" vocabulary than the Kim Jong-un regime's textbooks (see Tables 5 to 7). This result is in line with various earlier findings, which also reported differences in vocabulary difficulty levels between the two regimes. The second one indicated that the Kim Jong-un regime's Grade 1 textbook may contain more vocabulary words from the BNC/COCA 3K list than the Kim Jong-il regime's textbook. This result implies that Kim Jong-un's Grade 1 textbook may hold more challenging words than that of Kim Jong-il, suggesting a potential shift in the lexical complexity within Kim Jong-un's Grade 1 curriculum. These findings could indicate a broader trend in the transformation of English language education under Kim Jong-un's regime, with a potential focus on introducing more challenging words from BNC/COCA core vocabulary at earlier stages of learning while also mitigating the issue of gaps in difficulty levels between grades; it seems that they have fine-tuned the curricula to ensure the difference in difficulty levels between grades is even.

It is essential to match the vocabulary complexity of reading materials with the language proficiency of L2 learners, as this significantly impacts their enjoyment and success in learning a foreign language (Lewis, 1997; Schulz, 1981). Although previous studies have shown inconsistencies regarding the ideal token coverage threshold (e.g., Coxhead, Nation, & Sim, 2015; González-Fernández & Schmitt, 2020; Nation, 2006), at least 95% lexical coverage is commonly considered to be required for sufficient comprehension (Herman & Leeser, 2022; Hirsh & Nation, 1992; Laufer, 1989; Lee & Kim, 2022, 2023; Nation, 1990). Furthermore, to enjoy the reading experience without additional support, L2 learners should be familiar with over 98% of the text's vocabulary (Hu & Nation, 2000). Therefore, it is crucial to provide L2 learners with texts that meet or surpass the token coverage threshold, thereby allowing them to develop their L2 proficiency skills through reading.

This study's findings, based on the 95% lexical threshold level, have revealed that the textbooks from neither regime fully meet the universal vocabulary threshold standards, as suggested by Nation (2006) and Webb and Nation (2017). This indicates that North Korean middle school English textbooks may not be as effective in equipping students with the necessary vocabulary for language proficiency. However, the qualitative analysis revealed

that the Kim Jong-un regime's textbooks seem to be more accessible than the Kim Jong-il regime's ones, owing to the active use of foreign loan words in Korean. A noticeable difference in the lexical threshold level was observed between the Grade 2 and Grade 3 textbooks of both regimes. In both cases, the textbooks under Kim Jong-un's regime satisfied the lexical coverage threshold at lower BNC/COCA core lists, indicating that the revised textbooks are more accessible in terms of vocabulary quantity and variety.

When comparing the lexical coverage of North Korean middle school English textbooks with their South Korean counterparts, it is evident that there is a discrepancy in the level of language input. Lee and Kim (2023) found that South Korean middle school Grade 3 English textbooks were based on the BNC/COCA 2K~3K level. In contrast to the standards found in South Korea, the current findings indicate a potential disparity in vocabulary exposure and learning between North Korean and South Korean students. Addressing this discrepancy is crucial for ensuring a more balanced and equitable English language education for students in both countries; this, in turn, could contribute to fostering better communication and mutual understanding between people from North Korea and South Korea. For example, by examining the differences in lexical coverage and adjusting the curriculum accordingly, educators and policymakers can work toward providing a more effective and engaging learning experience for students across the Korean peninsula.

5.2. Lexical Features: A Comparison of Diversity, Readability, and Complexity

Our analysis revealed several differences between the English textbooks from the Kim Jong-il and Kim Jong-un regimes in terms of lexical features, such as token, type, STTR, MWL, and MSL. These findings are consistent with the literature review, highlighting the potential for lexical feature changes in response to educational reforms and revisions (see Lee & Kim, 2022; Lee, 2020). With the full implementation of the 2013 New Educational Guidelines and the Overall 12-Year Compulsory Education Guidelines, the Kim Jong-un regime has displayed an amicable attitude toward English education in general, in stark contrast to the Kim Jong-il regime. Consequently, English textbooks have been updated to meet global standards, reflecting the Kim Jong-un regime's recognition of English as a global language and its desire to enhance the country's international presence by boosting the quality and quantity of foreign language input, particularly at the lexical level.

Regarding tokens and types, this study found statistically significant differences across all grades when comparing English textbooks from the two regimes. The average number of tokens in the textbooks from the Kim Jong-il regime exhibited an inconsistent pattern: 31.3 in Grade 1, 30.9 in Grade 2, and 68.7 in Grade 3. However, the numbers increased progressively in the textbooks from the Kim Jong-un regime: 17.5 in Grade 1, 34.1 in Grade

2, and 46.6 in Grade 3. This suggests that the Kim Jong-un regime's textbooks have been designed in a more systematic manner, with the gaps across grades being more evenly distributed. The number of vocabulary types exhibited a similar trend. These changes suggest that the revisions made to the textbooks under Kim Jong-un's regime were aimed at better calibrating the educational materials to foster a more effective and engaging student learning experience. As Lee (2020) noted, the revised textbooks incorporate more authentic materials and examples with rich contextual and linguistic variety, which contribute to creating a more engaging and effective learning experience for students.

The analysis of STTR values revealed higher lexical diversity in the Kim Jong-un regime's textbooks compared to those of the Kim Jong-il regime, particularly for Grades 2 and 3. The increased lexical diversity may contribute to a more enriching and engaging learning experience for North Korean L2 learners of English, offering them a broader range of vocabulary. This observation is in line with the findings of Lee and Kim (2022), who reported higher lexical diversity in the Kim Jong-un regime's textbooks compared to the Kim Jong-il regime's textbooks.

Regarding MWL and MSL, it was found that the Kim Jong-un regime's textbooks contained increased word length across all grade levels as well as effectively adjusted sentence length differences across grade levels. The changes indicate that the language input in the Kim Jong-un regime's textbooks has been improved both quantitatively and qualitatively compared to the Kim Jong-il regime's textbooks. This has been achieved by incorporating a larger number of diverse vocabulary words, which represent various cultural and technological aspects; furthermore, the textbooks are suitably balanced across grades. This indicates the development of a more comprehensive and engaging educational experience for L2 learners.

In sum, this study provides valuable insights into the lexical coverage and features of North Korean middle school English textbooks under the Kim Jong-il and Kim Jong-un regimes. While no significant differences were found in the lexical coverage of textbooks between the two regimes, except for Grade 1, the analysis of lexical features revealed notable changes in diversity, readability, and complexity. These findings contribute to the existing body of research on English language education in North Korea by illuminating the potential implications of the differences in vocabulary exposure and lexical features between North Korean and South Korean students. The current findings highlight the importance of considering the impact of these disparities on the English language learning experiences of North Korean students, as well as the potential linguistic and cultural challenges they may face when interacting with their South Korean counterparts.

6. SUMMARY AND IMPLICATIONS

This study presents a comprehensive corpus-based analysis of the differences in English textbooks between the Kim Jong-il and Kim Jong-un regimes in North Korea. The primary objective is to explore the disparities in lexical coverage and features in North Korean middle school English textbooks under the two regimes and thus gain insights into the complexity, vocabulary, and readability of the learning materials. By analyzing BNC/COCA core baseword lists and lexical features, such as lexical diversity, MWL, and MSL, this study reveals the distinctions between the English textbooks used under the two regimes.

The results indicate that the Kim Jong-un regime has implemented reforms to enhance English language education by increasing lexical diversity, textual complexity, and vocabulary exposure. According to the findings, although there were no significant differences in core vocabulary distribution between the two regimes, there were interesting disparities in the lexical coverage (with the exception of Grade 1) and lexical features of the textbooks. The Kim Jong-un regime's textbooks show improvements in diversity, readability, and complexity, as evidenced by higher lexical diversity, longer average word length, and effectively adjusted MSL, in addition to a greater number of tokens and types. These changes highlight a more comprehensive and engaging learning experience, with improved language input in terms of both quantity and quality for North Korean students of English.

While further research is necessary to better understand the impact of educational reforms on the linguistic aspects of North Korean English textbooks, the current findings emphasize the need for policymakers and educators to take this study's insights into account when revising and developing future English language curricula. This would help ensure that North Korean students, as L2 learners of English, receive the language input required to become proficient in English and better prepare for interactions with their South Korean counterparts and the global community at large.

Although the quantitative analyses in this study did not reveal statistically significant differences in all instances between the two regimes, it is essential to acknowledge that these results are neither comprehensive nor absolute. Consequently, further qualitative exploration is required to better comprehend the linguistic nuances of North Korean English textbooks and discern any challenging or ideological vocabulary across BNC/COCA 1K to 34K baseword lists. Such additional analyses will facilitate a more holistic understanding of the linguistic differences and commonalities between these two North Korean regimes.

In light of international educational practice, the insights derived from our analysis extend beyond the North Korean context and have tangible implications for international educational practice. Educational reformers, curriculum designers, and language policymakers may draw upon our findings to navigate the intricate interplay between

political ideologies and language education. This discussion aims to shed light on how pedagogical strategies and curriculum development can be adapted to address political sensitivities in language instruction while promoting linguistic competence and critical thinking.

Our study underscores the need for policymakers to design language education policies that transcend mere linguistic proficiency. Effective policy must foster critical engagement with language materials, ensuring they serve as conduits for cultural exchange and political literacy. As such, the evolution of English education in North Korea serves as a case study for policy-driven educational reform in diverse socio-political landscapes.

Our corpus-driven exploration, while firmly rooted in the North Korean context, contributes to a broader discourse on the role of political ideology in language education. It provides a practical blueprint for international educators and policymakers who are committed to developing resilient and reflective language education curricula that can be applied worldwide.

Applicable level: Secondary

REFERENCES

Ahn, H., & Lee, H. (2022). Unveiling revised North Korean English textbooks: Language, ideology, and internationalisation. *3L The Southeast Asian Journal of English Language Studies, 28*(3), 17-34.

Anthony, L. (2020). AntWordProfiler (Version 3.5.9) [Computer software]. Laurence Anthony's website. Retrieved on March 21, 2024, from https://www.laurenceanthony.net/software/antwordprofiler/

Biber, D. (1991). *Variation across speech and writing*. Cambridge, England: Cambridge University Press.

Cho, J. (2014). Direction of North Korean education policy and reform of secondary education curriculum in the Kim Jong Un era. *Unification Policy Research (KINU)*, *23*(2), 177-206.

Choi, G. (2018). North Korean refugees in South Korea: Change and challenge in settlement support policy. *The Korean Journal of International Studies*, *16*(1), 77-98.

Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies*, *50*, 121-135.

Crossley, S. A., Allen, D., & McNamara, D. S. (2011). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, *16*(1), 89–108.

Davies, M. (2008). Word frequency data from the corpus of contemporary American English. English-corpora: COCA. Retrieved on May 15, 2024, from https://www.english-corpora.org/coca/

González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, *41*(4), 481-505.

Herman, E., & Leeser, M. J. (2022). The relationship between lexical coverage and type of reading comprehension in beginning L2 Spanish learners. *The Modern Language Journal*, *106*(1), 284-305.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689-696.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Hwang, S.-Y., & Kim, J. (2017). Comparison of the vocabulary before and after 2013 revision of North Korean English textbooks for junior middle schools. *The Journal of Learner-Centered Curriculum and Instruction*, *20*(5), 611-634.

Kang, H. S. (2020). Changes in English language policy in Kim Jong-un's North Korea: A prelude to reconciliation? *English Today*, *36*(1), 30-36.

Kim, J. (2020). Vocabulary difference of South and North Korean English textbook. *Journal of the Korea Contents Association*, *20*(1), 107-116.

Kim, J., Lee, J. Y., & Kim, J. R. (2017). A diachronic lexical analysis of the North Korean English textbooks. *The Journal of the Korea Contents Association, 17*(4), 331-341.

Kim, T.-Y. (2021). *Historical development of English learning motivation research: Cases of Korea and its neighboring countries in East Asia*. Cham, Switzerland: Springer.

KINU. (2021). *Online series CO21-04: Analysis on North Korea's 4th plenary meeting of the 8th central committee of the workers' party of Korea*. Seoul: Korea Institute for National Unification.

Kirkpatrick, A., & Liddicoat, A. J. (Eds.). (2019). *The Routledge international handbook of language education policy in Asia*. London: Routledge.

KIUE. (2021). *KIUE. 2020. Understanding North Korea.* National Institute for Unification Education. Seoul:Ministry of Education.

Lankov, A. (2014). *North of the DMZ: Essays on daily life in North Korea*. Jefferson, NC: McFarland.

Lankov, A. (2015). *The real North Korea: Life and politics in the failed Stalinist utopia*. Oxford, England: Oxford University Press.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Clevedon, England: Multilingual Matters.

Lee, K.-A. (2020). Analysis of changes in English textbooks in North Korea in the Kim Jong-un era. *Journal of Peace and Unification, 10*(3), 83-110.

Lee, S. (2014). A critical review of research on North Korean English textbooks: Focusing on the studies published in South Korea during 1994-2013. *The Journal of Foreign Studies*, *28*, 83-109.

Lee, Y. C. (2018). The hallmarks of L2 writing viewed through the prism of translation universals. *Linguistic Research*, *35*, 171-205.

Lee, Y. C. (2019). Spotting non-nativeness in L2 texts: A statistical approach to translationese. *Studies in English Language and Literature*, *45*(1), 367-388.

Lee, Y. C. (2021). Function words as markers of translationese: A corpus-based approach to mental translation in second language writing. *Korean Journal of English Language and Linguistics*, *21*, 261-281.

Lee, Y. C., & Kim, T.-Y. (2022). The paradigm shift in English language teaching in North Korea: A corpus-assisted analysis. *Korea Journal of English Language and Linguistics*, *22*, 279-299.

Lee, Y. C., & Kim, T.-Y. (2023). Elite English education in North Korea: A multifaceted corpus-based comparative analysis of English textbooks. *Korea Journal of English Language and Linguistics*, *23*, 554-570.

Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice*. Hove, England: Language Teaching Publications.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal, 96*(2), 190-208.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. London: Palgrave Macmillan.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381-392.

Nation, P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-82.

Nation, P. (2009). *Teaching ESL/EFL reading and writing*. London: Routledge.

Nation, P. (2013). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.

Nation, P. (2016). *Making and using word lists for language learning and testing*. Amsterdam, The Netherlands: John Benjamins.

Nation, P. (2017). The BNC/COCA Level 6-word family lists (Vers6-word0.0) [Data file]. Paul Nation's Resources. Retrieved on May 15, 2024, from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists

Schulz, R. A. (1981). Literature and readability: Bridging the gap. *Language Learning*, *31*(2), 319-332.

Scott, M. (2016). WordSmith Tools (Version 7.0) [Computer software]. *Lexical Analysis Software*. WordSmith Tools. Retrieved on May 15, 2024, from https://www.lexically.net/wordsmith/

Song, J. J. (2002). The Juche ideology: English in North Korea. *English Today*, *18*(1), 47-56.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford, England: Oxford University Press.

Yoo, H., & Kim, J. (2018). A comparison of structural organization of English textbooks between pre and post North Korean 2013 curriculum revision. *The Journal of the Korea Contents Association*, *18*(7), 412-422.