KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

Elite English Education in North Korea: A Multifaceted Corpus-based Comparative Analysis of English Textbooks*

Younghee Cheri Lee (Ajou University) **Tae-Young Kim** (Chung-Ang University)

Younghee Cheri Lee (1st author)
Teaching Professor, Dasan
University College, Ajou Univ.
Tel: 031-219-3053
Email: cheriberry@ajou.ac.kr

Tae-Young Kim
(corresponding author)
Professor, Dept. of English
Education, Chung-Ang Univ.
Tel: 02-820-5392
Email: tykim@cau.ac.kr

ABSTRACT

Lee, Younghee Cheri and Tae-Young Kim. 2023. Elite English education in North Korea: A multifaceted corpus-based comparative analysis of English textbooks. *Korean Journal of English Language and Linguistics* 23, 554-570.

This study offers a systematic investigation into the attributes of English language instruction in North Korea's elite high schools, utilizing a comprehensive, corpus-based method. We scrutinize various linguistic features—lexical coverage, lexical variety, lexical complexity, and syntactic complexity—that define the privileged education imparted under Kim Jong-un's regime. The findings unearth unique traits that distinguish English education in North Korean elite schools from general high school English education in South Korea. The study exposes a limited lexical coverage and a diminished lexical variety in North Korean instructional materials. Intriguingly, North Korean resources exhibit low lexical complexity despite a pronounced focus on scientific texts. Additionally, the mean sentence length, indicative of syntactic complexity, was markedly shorter than in South Korean curricula. These findings hint at potential constraints in providing North Korean gifted students adequate exposure to intricate syntactic structures, which may, in turn, limit their proficiency in engaging with advanced English texts in future academic or professional contexts. The paper culminates with examining the implications of these results, speculating on the factors contributing to such phenomena in the context of North Korean elite education compared to South Korean regular high schools. It advocates for including a more diverse set of lexical items and sentence structures in the curriculum to enrich the learning experience of North Korea's advanced students.

KEYWORDS

North Korea, elite education, gifted education, English education, English textbook, corpus

1. Introduction

Since the 1980s, under Kim Il-sung's leadership, North Korea has begun implementing a distinct educational strategy. This strategy identifies and nurtures extraordinary students through specialized education programs. Notable institutions embodying this approach include the Je-il (meaning: First) Middle School—an elite establishment—and Pyongyang's Man Kyung Dae Revolutionary School's national language institute. These institutions primarily cultivate talented students, addressing issues arising from economic stagnation and, more crucially, developing gifted individuals in science and technology (KINU 2021).

Established in Pyongyang in 1984 by Kim Il-sung's directive, the Je-il elite school was designed to educate students with exceptional talent and aptitude selectively and systematically. This concept swiftly expanded across North Korea, establishing similar institutions nationwide. Under the rule of Kim Jong-il in the 1990s, the focus on elite education intensified and carried into the tenure of Kim Jong-un. The sweeping educational reform of 2012, peaking with a curriculum overhaul in 2013, prompted substantial changes to the English textbooks utilized in elite high schools. Reports suggest that institutions like Je-il employ specialized textbooks akin to those found in higher education (i.e., junior or technical college), encompassing subjects such as natural sciences, computer science, and English (KIUE 2022).

Various research efforts have been conducted to analyze North Korea's education policies and instructional materials across three distinct regimes: Kim Il-sung, Kim Jong-il, and Kim Jong-un. Much of this research has concentrated on comparing English textbooks across these regimes, conducting comprehensive studies on North Korean English textbooks, ranging from textbook analysis to in-depth text examinations using a corpus method. However, most of these studies have been confined to general school levels, indicating a gap in the comparative study of general and elite schools.

Previous research findings indicate that the comprehensive English textbook reform following Kim Jong-un's ascendancy extended to public junior and senior high schools (comparable to middle and high schools in South Korea) and elite institutions like the Je-il Middle School. Regrettably, the dearth of studies concerning this specialized education and the limited comparisons between public and elite schools and between North Korean elite education and South Korean public education hamper understanding the full impact of the comprehensive 2013 curriculum reform on the scope and aims of elite education.

Furthermore, the need for more diverse analytical parameters and the absence of comprehensive explanations regarding what specific parameter results imply for North Korean elite education complicate the research process. To bridge these gaps, this study aims to apply a broad range of linguistic analysis indicators to North Korean elite education, specifically English education in high schools, and contrast them with South Korean general public education. This comparison aims to gauge the linguistic level and objectives pursued by Kim Jong-un's specialized education policy. To this end, the following research question was formulated:

(1) How do linguistic aspects—lexical coverage, lexical variety, lexical complexity, and syntactic complexity—differ between elite high school English in North Korea and regular high school English in South Korea?

(2) What insights arise from comparing North Korean elite English education and South Korean regular English education?

2. Recent Innovations of English Education and Elite Education in North Korea

In recent years, English language education is at the heart of educational reforms, particularly within school textbooks in North Korea. Recognizing the importance of English as a global lingua franca, North Korea has made significant strides in improving the quality of English education. The reform of the English curriculum in 2013, under Kim Jong-un's regime, was one such significant stride (see Lee 2020, Oh and Kim 2020). The administration of Kim Jong-il exhibited a shift in the objective of foreign language education; rather than preparing for warfare, the focus tilted toward facilitating interactions with overseas nations for the advancement of science and technology. This shift was substantiated when the Kim Jong-il regime established a strategic alliance with the British government to improve English education in North Korea following the formal establishment of diplomatic relations with the U.K. in December 2000. Supported by the U.K. government, the British Council supplied native English teachers, kick-starting an English education program aimed primarily at North Korea's elite in major Pyongyang universities. This collaboration extended beyond classroom instruction, encompassing overall improvements to North Korean English education, such as curriculum development, textbook writing, and teacher training.

Several prior findings claimed that under the educational regulations of North Korea's 2013 academic year, the number of English instruction hours allocated for junior high schools, which amounted to 408 hours, significantly surpassed those dedicated to 'revolutionary activities' (Cho 2014, Lee and Kim 2022). Similarly, in senior high schools, the 243 hours devoted to English instruction remained notably higher than those assigned to ideological education, including 'revolutionary activities.'

As outlined above, this curriculum for 2013, created under Kim Jong-un's reign to sustain the regime, reflects an interesting shift in priorities. The regime emphasizes English instruction more than traditionally required ideological education courses. Furthermore, the decision to phase out the Russian language course and streamline foreign language instruction into English exclusively is a significant change. This move vividly demonstrates the regime's intent to enhance its focus on English education since Kim Jong-un's rise to power.

Speaking of educational reforms, gifted education in North Korea was no different. As part of the broader educational apparatus, the gifted education system in North Korea has undergone a series of transformations reflective of the prevailing political leadership. The inception of gifted education traces back to the reign of Kim Il-sung in the 1980s, as this type of education was envisioned as a conduit to identify and cultivate individuals of exceptional talent and potential (KIUE 2021). This emphasis on cultivating exceptional talent was anchored in the belief that it was essential to mitigate various issues spurred by economic stagnation and foster prodigies in science and technology (KIUE 2022).

Under Kim Jong-il, these efforts were intensified in the 1990s. The gifted education system was consolidated and enhanced, focusing on expanding its reach and refining its curricula (KINU 2021). Educational institutions dedicated to gifted education, such as the Je-il Middle School and the Korean Language Academy in Pyongyang, were further developed during this period, reflecting the regime's commitment to fostering talent. With Kim Jong-un's ascent to power, the educational curriculum was overhauled in 2013 (KINU 2021). This change included a significant revamp of English textbooks used for elite education, demonstrating a clear commitment to globalizing and improving language proficiency among gifted students by including four language skills not confined to written language—reading and writing. Notably, these reforms were comprehensive, encompassing elementary, middle, and high schools and institutions that catered to gifted education, underscoring the regime's ambition to enhance education standards at all levels. The high standard of these textbooks suggests a commitment to equipping gifted students with advanced language skills.

Educational reform in North Korea, particularly regarding English education, has been an ongoing area of focus in domestic academia, with numerous studies conducted to date. The primary focus of these studies has been on English textbooks in North Korea, often involving comparative analyses between textbooks revised during different leadership periods and across varying levels of schooling, such as elementary, middle, and high schools (e.g., Park and Shin 2016). Recently, research has shifted toward studying the linguistic features in English textbooks used in gifted education.

For instance, Kim and Hwang (2018) constructed a corpus of English textbooks from Je-il Middle School, an institution dedicated to nurturing gifted students in North Korea. They compared them with English textbooks from South Korean and North Korean general middle schools. Their findings suggested that the English textbooks from the Je-il Middle School contained a higher incidence of British spelling, vocabulary indicating parts of speech, and ideological vocabulary than South Korean middle school English textbooks. They also reported that though the variety of vocabulary was less in the North Korean textbooks, the difference was not substantial. However, the study has limitations as the data analyzed were based on versions published before the 2013 educational reform, and the textbooks compared were also from the pre-Kim Jong-un era. The study only compared a few grades of middle school, and a lack of appropriate inferential statistical verification made the generalization of the results challenging. Beyond this, while a few other studies address gifted education, they mainly discuss it from a cultural perspective, leaving the study of English as represented in texts used in gifted education relatively unexplored.

Kim (2020), while not directly investigating gifted education materials, compared collocational patterns between the corpus of English textbooks from North Korean and South Korean high schools using N-grams. The analysis revealed a higher relative prevalence of noun phrase collocations in North Korean textbooks, irrespective of spoken or written language, due to the repetition of academic vocabulary and political or ideological terms.

Lee and Kim (2022) performed an in-depth corpus-based analysis utilizing various analytical indices related to the evolution of North Korean foreign language education policies and the history of North Korean English textbooks. They established a comparable corpus of English in North Korea for both Kim Jong-il and Kim Jong-un periods and compared changes in content, quantitative changes in language input, and qualitative changes in language input using measures like token coverage, STTR, concordance plot, and 3-gram lexical bundles. Their analysis suggested that the revisions made to the textbooks during Kim Jong-un's era met the lexical threshold (95%) of the BNC/COCA 3K core vocabulary, indicating that the revised versions were meticulously regulated with an optimal amount of language input. Concerning the quality of language input, the revised textbooks under Kim Jong-un's regime incorporated a more significant proportion of authentic and high-frequency 3-gram lexical bundles.

Despite these advancements and transformations in North Korea's gifted education, research on the English education provided at North Korean elite schools remains limited and largely imprecise. Existing research is mostly frequency-based or limited to a few linguistic parameters, providing only a partial understanding of elite English education. Furthermore, a considerable portion of the research is based on textbooks from Kim Jong-il, with limited access to the revised textbooks introduced under Kim Jong-un (KINU, 2021).[1] Consequently, there is a pressing need for more comprehensive and accurate research on English education in North Korean elite schools.

This research, thus, aims to address these gaps in the literature, exploring the linguistic and pedagogical aspects of English education in North Korean elite schools and comparing them with those in South Korean public high

---

[1] Only outdated textbooks from the Kim Jong-il era (published in 2007) had been accessible at the Ministry of Unification's North Korean Center. The revised textbooks from Kim Jong-un's tenure have been absent, creating a significant gap in research, though textbooks from elite senior high schools are available for inspection.

schools. By doing so, the study aims to provide a more nuanced understanding of the objectives, standards, and quality of English education in North Korean elite schools.

As highlighted in our previous research, we have noted a shift in North Korea under Kim Jong-un's rule toward presenting itself as a "normal" country. We are particularly interested in whether elite education reflects these trends. The choice of comparing North Korean elite English education with mainstream South Korean education is purposeful, as South Korea serves as a significant geopolitical and cultural reference point for North Korea. We believe it is essential to understand the context of North Korean elite education as it reflects the expectations of the country's leadership, which can provide unique insights into the educational aspirations and future directions of North Korea. The focus on North and South Korea is based on their shared cultural and historical ties, unique political division, and contrasting educational philosophies, which make the comparison particularly revealing and significant. We believe this comparative study can provide a better understanding of the dynamics of English education in North Korea and contribute to the broader discourse on English language education in politically contrasting contexts.

3. Methods

3.1 Corpus Data

This research aims to explore the intricate aspects of elite English education in North Korea during Kim Jong-un's regime. Its objective is to uncover the strategies and policies adopted within North Korea's elite education system by comparing it with the mainstream English education provided at equivalent grade levels in South Korea. The decision to compare North Korean elite English education with South Korean mainstream English education is based on the premise that such a comparison can offer valuable insights into the standards that North Korean authorities aspire to set for their top-tier students. Given that the South Korean education system is globally acknowledged for its proficiency in English, it serves as a suitable benchmark. Another aim of this research is to determine whether the English education intensity offered to elite students in North Korea matches the standard level found in South Korea. The results of this comparison can provide a greater understanding of the implications of such educational strategies in North Korea's wider sociopolitical context.

To facilitate this investigation, a corpus-based quantitative text analysis is conducted on English textbooks, which serve as the main instructional resources. The corpus created for this purpose includes English study materials used in Kim Jong-un's elite North Korean high schools as well as regular South Korean high schools. The corpus encompasses resources from three grades and spans from the 2010s to the present.

To compile this corpus, we obtained the revised textbooks intended for elite senior high schools in North Korea during Kim Jong-un's regime. These resources were accessed through the Ministry of Unification's North Korean Center. The methodological framework for this investigation draws from the work of Lee and Kim (2022). The primary focus is on the key reading sections within each grade's English textbooks, excluding directive statements. The analysis is concentrated on textbooks for grades 1, 2, and 3 from elite high schools restructured under Kim Jong-un's leadership. In South Korea, we analyzed a range of textbooks under public education: ten different government-approved textbooks for the 1st grade and nine for the 2nd grade.[2] The primary objective of our

---

[2] The first-year English textbook corpus in South Korea includes works from Neungyule authored by Kim Sung-gon and Yang Hyun-Kwon, YBM by Park Joon-Eon and Han Sang-ho, Kumsung by Choi In-chull, Dong-a by Lee Byung-min, Visang

analysis is to discern the distinguishing characteristics of North Korean elite English education in contrast to the regular English education provided in South Korea. As such, we defined North Korean elite English as the "study corpus," with South Korean regular English serving as the "reference corpus." According to established principles within corpus linguistics, the ideal size of a reference corpus is typically considered to be at least five times larger than the size of the study corpus.

Regarding the discrepancy in the sizes of the corpora we used in our study, it might seem as though this discrepancy could impact the reliability and accuracy of our results. However, it is essential to clarify a few points regarding corpus size in linguistic research. First of all, following guidelines from previous corpus linguistics studies, the size of our reference corpus (South Korean textbooks) was intentionally larger, approximately five times the size of our study corpus (North Korean textbooks). This is a well-accepted practice in corpus linguistics designed to ensure that the reference corpus provides a robust linguistic environment in which to compare and analyze the study corpus. Second, the sizes of the corpora in studies of this kind are often determined by the availability of data. In our case, the North Korean corpus is inherently limited due to the scarcity of accessible material. The comparison with a larger South Korean corpus still holds validity, considering it provides a broader base for reliable and meaningful comparative analysis. Finally, although corpus size may influence the frequency of certain linguistic features, our analysis focused on proportional measures (e.g., standardized type-token ratio, mean word length, etc.) which are not impacted by corpus size. We also used normalization procedures to control size differences when comparing absolute frequencies. The comprehensive breadth of the comparable corpora is delineated in Table 1.

Table 1. Corpus Design and Textual Statistics

Group	Sub-Corpus	Equivalent to South K. Grade	Token (#)	Type (#)	Sentence (#)
North Korea (NK): Kim Jong-un's *Elite High (EH) School English*	NK_EH1	Elite High School: Grade 1	4,033	1,192	344
	NK_EH2	Elite High School: Grade 2	3,413	1,087	234
	NK_EH3	Elite High School: Grade 3	3,979	1,255	245
South Korea (SK): *High (H) School English*	SK_H1	High School: Grade 1	65,210	6,917	4,679
	SK_H2	High School: Grade 2	57,900	7,389	3,621
	SK_H3_CSAT	High School: Grade 3	6,981	2,173	299
	SK_H3_KICE	High School: Grade 3	19,878	4,164	877

As shown in Table 1, for the 3rd grade, high school students typically focus on materials for the College Scholastic Ability Test (CSAT) rather than a standard public textbook. Therefore, we incorporated two kinds of CSAT-related texts into our analysis. One category was the CSAT Mock Test administered by the Korea Institute for Curriculum and Evaluation (KICE) from 2018 to 2021, conducted twice annually, every June and September. The other category comprised the actual College Scholastic Ability Test during the same period. However, rather than including all questions from these tests, we selected only the most challenging question items dealing with "main idea, grammar, vocabulary, inference, and sequence." Through this comprehensive comparative analysis, our study aimed to shed light on the unique characteristics and stark differences between English education in

---

by Hong Min-pyo, Jihaksa by Min Chan-kyu, and Chunjae by Kim Tae-Young and Lee Jae-young. The second-year English textbook corpus includes works from Neungyule authored by Kim Sung-gon, YBM by Park Joon-Eon and Han Sang-ho, Kyohaksa by Kang Moon-goo, Kumsung by Choi In-chull, Dong-a by Kwon Hyuk-seung, Visang by Hong Min-pyo, Jihaksa by Min Chan-kyu, and Chunjae by Lee Jae-young.

North and South Korea.

To perform inferential statistics, we partitioned the corpus files into separate units and texts. This means that instead of treating each entire corpus as a single data point, we subdivided it into smaller data points. Each text within the corpus was treated as a distinct data unit, and each of these units was analyzed independently. This allowed us to gather multiple N values from each corpus, which provided a more detailed and nuanced analysis. This approach enabled us to take into account intra-corpus variability, that is, the linguistic differences within a single corpus. As a result, the method can give us a more precise understanding of the linguistic features under investigation.

3.2 Encoded Variables and Tools

In our quest to answer the above research questions, we conducted a text analysis using four parameters, each corresponding to a different linguistic perspective—lexical coverage, lexical variety, lexical complexity, and syntactic complexity. To initially gauge the level of elite education under Kim Jong-un's rule, we carried out a step-by-step analysis based on vocabulary lists derived from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), which served as our reference corpora (see Davies 2008, Nation 2016). The intention was to evaluate the "Englishness" of the language used in North Korean textbooks against objective criteria drawn from these two representative corpora of British and American English. We used the BNC/COCA baseword lists 1K to 34K (Nation 2006) to measure lexical coverage and analyze the difficulty level of the texts.

In our analysis, we utilized a vocabulary profiling program to evaluate the text corpus. The minimum lexical coverage threshold for comprehension was established at 95%. This means that to comprehend the texts adequately, readers need to be familiar with at least 95% of the words used. This threshold reflects a fundamental principle in language acquisition, often referred to as the "coverage hypothesis" (Hu and Nation 2000). It suggests that a high level of lexical understanding is needed to make sense of a text, with various studies indicating that 95% to 98% coverage is the minimum for fluent comprehension (Hu and Nation 2000, Laufer 1992, Schmitt, Jiang, and Grabe 2011).

For the second parameter, we then calculated the Standardized Type/Token Ratio (STTR) to assess the qualitative level of the texts based on lexical diversity. The STTR is an index that standardizes the ratio of word types to word tokens in a text, irrespective of its length. A higher STTR value indicates a broader variety of vocabulary and, by extension, a higher quality of text (see Malvern, Richards, Chipere, and Durán 2004). It has also been found that the closer a text is to the characteristics—i.e., nativeness—of native English, the higher the STTR index appears (Lee 2018, 2019, 2021, Lee and Kim 2022).

Noting the sensitivity of Type-Token Ratio (TTR) to text length addressed by Lu (2012), we adopted the Standardized Type-Token Ratio (STTR) as an indicator of lexical diversity in our study. The STTR is a variant of TTR specifically designed to mitigate the effects of text length on the measure. In our study, we used WordSmith Tools to calculate the STTR, setting the STTR basis of each sample size to 50-word tokens, which is a commonly adopted practice in corpus linguistics. This means that we calculated the type-token ratio for each successive sample of 50 words in the corpus, then averaged these values. By doing so, we ensured that the measures of lexical diversity were not artificially inflated or deflated due to variations in text length between the two corpora.

Adding to our investigative tools, we also utilized the Mean Word Length (MWL) as an indicator, which furnished a lens to examine the complexity of the employed lexicon. Fundamentally essential in linguistic studies, readability assessments, and text analyses, MWL delivers valuable insights by determining the average length of

words contained within a specific language sample or textual composition (Lu 2012). The MWL was chosen as a representative measure in our study primarily due to its relatively straightforward interpretability and wide usage in previous research as an indicator of lexical complexity. It has been observed that longer words often correlate with higher-level vocabulary and thus a greater level of complexity. In our study, MWL was calculated as the total number of characters (excluding spaces and punctuation) divided by the total number of words in the corpus. This calculation yields the average length of a word in terms of characters.

Our final analytical tool was the Mean Sentence Length (MSL), employed to gauge the complexity at a syntactic level. As learners' language proficiency increases, their learning materials typically display an increase in MSL, thus acquainting them with more complex linguistic structures and concepts. In this study, MSL was computed as the total number of words divided by the total number of sentences in the corpus. This gives the average length of a sentence in terms of words. It is important to note that the MSL can be influenced by the context and content imbued within the texts.

Text preprocessing for corpus data construction was performed using WordSmith Tools 7.0 (Scott 2016). The lexical coverage analysis based on BNC/COCA baseword lists was conducted using the Vocabulary Profile feature in AntWordProfiler 1.5.1w (Anthony 2021). Analysis of lexical variety, lexical complexity, and syntactic complexity was executed using WordSmith Tools 7.0 (Scott 2016).

To evaluate the statistical significance between the North and South Korean corpora, we used the *F*-Test Two-Sample for Variances as the tool for normality test analysis. In contrast, *t*-test functions were used to handle the assumption tests of heteroscedasticity/equal variance between the two groups. Regarding the risk of Type I errors associated with multiple comparisons and the option to apply the Bonferroni correction, we acknowledge that performing multiple comparisons without adjustment can potentially inflate the risk of Type I error. However, it is also important to note that Bonferroni corrections are highly conservative and can substantially increase the risk of Type II errors, potentially obscuring meaningful findings. In our study, the number of comparisons made was relatively moderate and based on a clear theoretical framework. Moreover, the *F*-tests we conducted had a specific and predetermined direction, reducing the risk of spurious findings. We also maintained a stringent *p*-value threshold, further mitigating this risk.

The statistical program we employed was the Analysis ToolPak built into Microsoft 365. The overarching objective behind these measures was to meticulously observe the goals pursued by elite English education under Kim Jong-un's regime through an empirically grounded lens.

4. Results and Discussion

4.1 Lexical Coverage

We turned our scrutiny toward text coverage to grasp the level of English instruction permeating elite education under the Kim Jong-un regime and discern how it diverges from public education in South Korea. The analysis unveiled outcomes starkly inconsistent with our initial expectations across all grades.

For the first grade, as indicated in Table 2, the English textbook for the elites in Kim Jong-un's high schools satisfied 96.45% of BNC/COCA 2K. In contrast, ten types of English textbooks for the first grade of regular high schools in South Korea satisfied 96.32% of BNC/COCA 3K. The distinct variance in vocabulary types was particularly striking when examined against the BNC/COCA 3K standard. The elite textbooks under Kim Jong-un covered 40 unique vocabulary types, while the regular South Korean textbooks encompassed 904, underscoring a

substantial disparity between the two groups.

This pattern remained broadly consistent in the second grade. As shown in Table 2, elite English textbooks in Kim Jong-un's high schools satisfied 95.98% at BNC/COCA 2K, just as in the first grade. On the other hand, the nine types of general English textbooks in South Korea achieved 95.38% at BNC/COCA 3K, once again revealing a significant gap between North Korean elite English and South Korean general English. Scrutiny of vocabulary types based on BNC/COCA 3K revealed that the elite North Korean textbook incorporated only 66 unique vocabulary types at this level. In contrast, its South Korean counterpart included over a thousand.

Table 2. Lexical Coverage Comparison: High School Grade 1

	BNC/COCA 1K~34K	North Korea: *Elite High School*		BNC/COCA 1K~34K	South Korea: *High School*	
Level	Baseword List	Cum. Token (%)	Type (#)	Baseword List	Cum. Token (%)	Type (#)
	31K_Proper Noun	5.09	60	31K_Proper Noun	1.80	366
	32K_Marginal	5.55	3	32K_Marginal	1.94	17
	33K_Compound	6.06	16	33K_Compound	2.37	113
	34K_Acronym	6.59	5	34K_Acronym	2.41	13
1	1K	91.00	732	1K	84.72	2,413
2	2K	96.45	155	2K	92.90	1,532
3	3K	97.73	40	3K	96.32	904
4	4K	98.74	26	4K	97.58	380

Table 3. Lexical Coverage Comparison: High School Grade 2

	BNC/COCA 1K~34K	North Korea: *Elite High School*		BNC/COCA 1K~34K	South Korea: *High School*	
Level	Baseword List	Cum. Token (%)	Type (#)	Baseword List	Cum. Token (%)	Type (#)
	31K_Proper Noun	4.14	51	31K_Proper Noun	1.86	356
	32K_Marginal	4.51	2	32K_Marginal	1.95	15
	33K_Compound	4.96	13	33K_Compound	2.31	103
	34K_Acronym	5.33	2	34K_Acronym	2.36	10
1	1K	88.94	675	1K	81.26	2,399
2	2K	95.98	166	2K	90.81	1,664
3	3K	98.51	66	3K	95.38	1,090
4	4K	99.16	20	4K	97.11	463

The third-grade scenario was inspected in two contexts: CSAT and KICE. As illustrated in Table 4, the vocabulary breadth of Kim Jong-un's third-grade elites reached 97.6% at the BNC/COCA 3K level, including 99 unique vocabulary types. In contrast, the CSAT corpus achieved 94.76% with 407 unique vocabulary types at the same level, and 96.74% with 99 types at BNC/COCA 4K. This comparison reveals that even the final stage of elite education in Kim Jong-un's regime does not quite meet the standards held by typical third-grade South Korean students preparing for university entrance exams.

Upon comparison with the corpus of the mock CSAT examination implemented by South Korea's KICE, the English proficiency of Kim Jong-un's third-grade elites attained 97.6% at BNC/COCA 3K and encompassed 99 vocabulary types. Conversely, the KICE corpus achieved 95.43% at the same level and incorporated 847 unique vocabulary types (See Table 5).

Table 4. Lexical Coverage Comparison: High School Grade 3_CSAT

Level	BNC/COCA 1K~34K Baseword List	North Korea: *Elite High School* Cum. Token (%)	Type (#)	BNC/COCA 1K~34K Baseword List	South Korea: *High School* Cum. Token (%)	Type (#)
	31K_Proper Noun	2.92	42	31K_Proper Noun	0.73	36
	32K_Marginal	3.29	4	32K_Marginal	0.00	0
	33K_Compound	3.66	13	33K_Compound	1.00	16
	34K_Acronym	4.03	2	34K_Acronym	1.04	2
1	1K	84.71	757	1K	74.16	906
2	2K	94.38	223	2K	86.02	464
3	3K	97.60	99	3K	94.76	407
4	4K	98.69	36	4K	96.74	99

Table 5. Lexical Coverage Comparison: High School Grade 3_KICE

Level	BNC/COCA 1K~34K Baseword List	North Korea: *Elite High School* Cum. Token (%)	Type (#)	BNC/COCA 1K~34K Baseword List	South Korea: *High School* Cum. Token (%)	Type (#)
	31K_Proper Noun	2.92	42	31K_Proper Noun	0.92	96
	32K_Marginal	3.29	4	32K_Marginal	1.00	5
	33K_Compound	3.66	13	33K_Compound	1.38	46
	34K_Acronym	4.03	2	34K_Acronym	1.39	2
1	1K	84.71	757	1K	74.59	1442
2	2K	94.38	223	2K	86.86	895
3	3K	97.60	99	3K	95.43	847
4	4K	98.69	36	4K	97.44	258

Contrary to conventional expectations, the caliber of North Korean elite English education seems significantly lower than that of regular South Korean English education. Generally, regardless of the location, gifted education tends to aim for a higher level of input than regular education. Thus, these findings present an unexpected deviation from the norm. These findings challenge the effectiveness of the sweeping overhaul of English textbooks intended to elevate the quality of foreign language instruction to global standards following Kim Jong-un's ascent to power. They seemingly underscore a failure to control and fine-tune the adequate level of language input at the core of elite education. As Lee and Kim (2022) asserted, careful consideration of practical input settings and the vocabulary level forming the English textbook content is critical for successfully implementing education reform, especially in North Korean English classrooms where another educational infrastructure is limited. As a result, texts composed of appropriate-level vocabulary need to be used as learning materials to achieve the intended language instruction.

The failure to regulate text levels seems to stem from a lack of expertise equivalent to textbook authors and inadequate infrastructure to match Kim Jong-un's ambition for educational reform. Moreover, this lexical coverage reiterates the findings of Lee and Kim (2022), which centered on the analysis of public high school English textbooks in North Korea during the reign of Kim Jong-un. Their research reported an overall adjustment in the quantity and quality of language input, including incorporating high-frequency vocabulary and lowering difficulty across all grades compared to the Kim Jong-Il era.

4.2 Lexical Variety

One helpful tool for analyzing the variety of vocabulary within a text is the Standardized Type-Token Ratio (STTR). Higher STTR values imply a more decadent assortment of words in the text, while lower values denote a somewhat limited lexicon. Our analysis discovered substantial differences in STTR values when comparing the lexical variety of English taught in North Korean elite schools with that in South Korean public education, as demonstrated in Table 6.

Table 6. Lexical Variety Comparison

Group	Corpus	*F*-test			*t*-test				
		F	*p*		*N*	*M*	*VAR*	*t*	*p*
NK_EH1 SK_H1	Elite Textbook Textbook	3.278	0.000	$p < .05$	30 84	95.943 96.853	2.639 0.805	-2.915	*p = .003***
NK_EH2 SK_H2	Elite Textbook Textbook	3.588	0.000	$p < .05$	30 53	96.119 96.890	2.576 0.718	-2.444	*p = .009***
NK_EH3 SK_H3_CSAT	Elite Textbook CSAT	1.350	0.189	$p > .05$	30 40	96.117 96.755	2.241 1.660	-1.914	*p = .029**
NK_EH3 SK_H3_KICE	Elite Textbook KICE	0.999	0.522	$p > .05$	30 112	96.117 96.560	2.241 2.243	-1.439	*p = .076*

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

Particularly in high school textbooks, these differences in the STTR parameter were statistically significant. It is noteworthy that the English language instruction in the elite schools during Kim Jong-un's regime exhibited a lower lexical variety in the first and second grades of high school compared to the traditional public education in South Korea (NK_EH1 vs. SK_H1: $p = .003**$, NK_EH2 vs. SK_H2: $p = .009**$). Regarding high school seniors, the CSAT displayed marginal significance, while the average difference between the regimes in the case of KICE was not statistically established (NK_EH3 vs. SK_H3_CSAT: $p = .029*$, NK_EH3 vs. SK_H3_KICE: $p = .076$).

These findings mirror the patterns previously identified in lexical coverage. Compared to those used in South Korea's mainstream high schools, the reduced lexical variety found in the English textbooks used in North Korea's elite education under Kim Jong-un's regime might challenge the perceived authenticity—"nativeness" or "Englishness"—of the respective texts. The terms "nativeness" or "Englishness" here refer to the intrinsic linguistic properties that endow the language with its authentic English character. This encompasses not only correct grammar and vocabulary usage but also the incorporation of a wide range of vocabulary words, apt usage of idiomatic expressions, collocations, and cultural references typical of English-speaking communities. The absence of such "nativeness" or "Englishness" in North Korean textbooks could detrimentally impact the quality of language input, potentially making the language taught somewhat estranged from the practical use of English in real-world contexts. This could consequently hinder the learners' ability to communicate effectively in English beyond textbook or classroom confines.

It should be noted that textbooks under Kim Jong-un's regime seem to have expanded their focus on scientific topics. However, even if these advanced subjects introduce new vocabulary, a lack of sufficient consideration for the diversity of words, in relation to the qualitative aspects of language input in instructional materials, might be at odds with the fundamental goals of gifted education. This implies that education should not merely revolve around acquiring new words but also fostering exposure to a broad range of vocabulary used in linguistically native and contextually meaningful ways.

The current results underscore the imperative role of diverse, authentic, and complex vocabulary in English language instruction, particularly in settings aiming to provide advanced English proficiency. The pronounced disparity between the North Korean elite and South Korean mainstream education accentuates the importance of expanding the lexical scope to facilitate comprehensive learning. Furthermore, it emphasizes the need to invest in resources and expertise to bring educational content in line with global standards, which may currently be absent in North Korea's gifted educational context.

4.3 Lexical Complexity

We utilized Mean Word Length (MWL) as the third parameter in our comparative analysis of North and South Korean high school English textbooks. Generally, MWL is gauged by the count of characters in a word, where longer words render the text more challenging to read and understand. This measure is frequently used in a variety of readability indices, all of which use word length in their computations, such as the Automated Readability Index (ARI), Coleman-Liau Index, and LIX (Läsbarhetsindex, a Swedish readability measure).

The MWL was measured in North and South Korean high school English textbooks for a comparative analysis of lexical complexity, yielding the following results. As detailed in Table 7, a high level of statistical significance was observed between the two sub-corpora across all grade levels in terms of the MWL parameter. Notably, a marginal level of statistical significance was identified in the first grade (NK_EH1 vs. SK_H1: p = .016*), with the statistical significance increasing markedly in the subsequent grades (NK_EH1 vs. SK_H1: p = .016*, NK_EH2 vs. SK_H2: p < .001***, NK_EH3_CSAT vs. SK_H3_CSAT: p < .001***, NK_EH3_KICE vs. SK_H3_KICE: p < .001***). Intriguingly, for the first grade, the MWL in elite English textbooks under Kim Jong-un's regime was marginally shorter than in the ten types of government-approved South Korean high school textbooks, yielding marginal significance.

Table 7. Lexical Complexity Comparison

Group	Corpus	F-test			t-test				
		F	*p*		*N*	*M*	*VAR*	*t*	***p***
NK_EH1 SK_H1	Elite Textbook Textbook	1.756	0.025	p < .05	30 84	4.343 4.471	0.083 0.047	-2.212	***p*** = .016*
NK_EH2 SK_H2	Elite Textbook Textbook	1.234	0.250	p > .05	30 53	4.403 4.631	0.090 0.073	-3.556	***p*** < .001***
NK_EH3 SK_H3_CSAT	Elite Textbook CSAT	1.074	0.412	p > .05	30 40	4.617 5.123	0.124 0.116	-6.060	***p*** < .001***
NK_EH3 SK_H3_KICE	Elite Textbook KICE	1.096	0.356	p > .05	30 112	4.617 5.139	0.124 0.113	-7.472	***p*** < .001***

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

As previously mentioned, the discrepancies were highly significant from the second to the third grade. In particular, the MWL for the third grade surpassed 5.0 in both the South Korean CSAT and KICE mock exams, while it remained in the mid-4s in elite third-grade textbooks. The analysis additionally revealed that North Korean elite English for the second grade had a shorter MWL than South Korean first-grade regular high school English, and North Korean elite English for the third grade was shorter than South Korean second-grade high school English.

Echoing earlier findings from the STTR analysis on lexical variety, these results suggest that elite English under Kim Jong-un's regime does not offer a sufficient challenge regarding the quantity and quality of language input.

A longer MWL facilitates more profound vocabulary learning from the perspective of L2 learners. For instance, in English, words can be dissected into roots, prefixes, suffixes, as well as inflectional endings, offering varied instructional strategies for expanding vocabulary size. However, if the words are not sufficiently long, these more advanced learning opportunities may be unavailable, potentially obstructing students' progress. This point underscores MWL's significance in various readability indices such as ARI, Coleman-Liau Index, and LIX, and its importance in assessing the complexity of English textbooks. By providing elite students with longer, more intricate words, textbooks may offer more abundant opportunities for the depth and width of vocabulary learning in their privileged English curriculum.

4.4 Syntactic Complexity

As a final measure to understand the characteristics of English in elite North Korean high schools, we explored syntactic complexity. The mean sentence length (MSL) is a significant indicator of syntactic complexity. This metric calculates the average sentence length in a given text by dividing the word count by the total number of sentences. We employed MSL to compare the English of elite North Korean high schools with high school English in South Korea. The outcomes of our analysis are as follows.

As demonstrated in Table 8, statistical significance was evident across all grade levels (NK_EH1 vs. SK_H1: $p < .001***$, NK_EH2 vs. SK_H2: $p < .001***$, NK_EH3_CSAT vs. SK_H3_CSAT: $p < .001***$, NK_EH3_KICE vs. SK_H3_KICE: $p < .001***$). Notably, English for second-grade students under Kim Jong-un's regime showed a shorter average sentence length compared to South Korean first-grade general English. Similarly, the average sentence length of third-grade elite English was observed to be shorter than South Korean second-grade English.

Table 8. Syntactic Complexity Comparison

Group	Corpus	F-test			t-test				
		F	p		N	M	VAR	t	p
NK_EH1 SK_H1	Elite Textbook Textbook	1.412	0.114	$p > .05$	30 84	11.538 14.319	9.487 6.717	-4.795	$p < .001***$
NK_EH2 SK_H2	Elite Textbook Textbook	0.959	0.462	$p > .05$	30 53	13.599 16.304	7.387 7.700	-4.298	$p < .001***$
NK_EH3 SK_H3_CSAT	Elite Textbook CSAT	0.391	0.005	$p < .05$	30 40	15.085 23.899	6.989 17.865	-10.691	$p < .001***$
NK_EH3 SK_H3_KICE	Elite Textbook KICE	0.189	0.000	$p < .05$	30 112	15.085 23.743	6.989 36.903	-11.544	$p < .001***$

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

Note that the MSL in both South Korean third-grade CSAT and KICE surpasses 23 words, matching the level of English articles in The Korea Times or The Korea Herald (See Goh and Lee 2016). In contrast, the MSL in North Korean third-grade elite English was significantly lower, at around 15 words. According to research, the MSL for general text is usually about 15 words (Lee and Jwa 2023). However, such a general text may not sufficiently address specific professional subject areas. While a substantial portion of scientific texts was included in the elite third-grade English of North Korea, the actual sentence length appears to be relatively short. Texts from specialized fields not supported by sufficient length are likely to have monotonous sentence structures. In such cases, learning to discern subtle contextual differences may not be adequately achieved. The syntactic complexity, which influences learners' reading comprehension, is essential to the curriculum scope and sequence in gifted

education. Nevertheless, the results of our analysis seem to suggest that such a need is not adequately addressed in Kim Jong-un's privileged English curriculum.

Sentence length is crucial in evaluating text quality as it determines the number of words in a sentence. Longer sentences may employ more complex syntactic structures. For instance, when an adjective modifies a noun, lengthier sentences may use extended modifiers like adjective phrases, prepositional phrases, participle phrases, or relative clauses with relative pronouns instead of a single adjective word. Various conjunctions also extend the sentence length, implying that a longer MSL often leads to exposure to more complex syntactic structures. Hence, such parameters are usually employed when assessing text complexity or readability indices.

To foster a more prosperous learning environment, elite English textbooks in North Korea need to offer sentences of varied lengths and complexities. By doing so, they would provide gifted students with the necessary tools to navigate advanced readings in the future, such as professional articles or technical science reports that Kim Jong-un expects his people to excel in.

Moreover, considering the role of diverse syntactic structures in fostering language complexity, a greater emphasis on sentence construction in the learning materials is also crucial. As our study indicates, shorter sentences may limit students' exposure to complex structures, such as prepositional phrases, participle phrases, or relative clauses, which might hinder their ability to grasp more sophisticated English texts later in their academic or professional life. Therefore, English education should prioritize introducing varied sentence structures to students, particularly in elite schools. Such will provide them with a more holistic and practical understanding of the language, equipping them to handle a broader range of syntactic structures in their future encounters with English.

Overall, regarding the connection between linguistic complexity in North Korean English textbooks and the education policies of North Korea, it can be interpreted that North Korea's emphasis on lexical simplicity may mirror the focus of its educational policy on learner accessibility, international communication, and engagement. On the other hand, South Korea's more extensive use of varied vocabulary and complex sentence structures may reflect its emphasis on global competitiveness and comprehensive English proficiency.

5. Conclusion

This research offers a comprehensive, corpus-based examination of English education in North Korea's elite high schools. The findings provide a disconcerting glimpse into the constraints of English language instruction in the country, presenting evidence of limited lexical coverage, diminished lexical variety, low lexical complexity, and reduced syntactic complexity compared to South Korean general high school education.

While the North Korean regime, under Kim Jong-un's leadership, aimed to elevate English language education to global standards, our study's findings suggest otherwise. Despite the increased inclusion of scientific texts in the curriculum, the English instructional materials fall short of offering a diverse, complex, and authentic language input. This shortfall, interestingly, is particularly pronounced in the elite educational sector that is expected to produce the best English speakers in the country.

These shortcomings point to several potential challenges. The findings suggest that the quantity and quality of language input available to the students could be at par with what is required for effective language learning. A lack of lexical variety and complexity could limit students' vocabulary expansion and exposure to advanced linguistic structures. Additionally, shorter sentence lengths in the instructional materials may impede the development of a nuanced understanding of complex syntactic structures.

This study's implications extend beyond the bounds of English language education in North Korea. The results

provoke a broader reflection on the educational strategies employed in contexts with limited resources and expertise. They underscore the importance of adequate language input settings and the judicious selection of vocabulary and sentence structures in curriculum design. Such considerations are critical for successful education reform, particularly in environments seeking to raise language instruction to global standards.

Moreover, we argue that North Korean gifted education settings need to include a more diverse set of lexical items and sentence structures in the curriculum to enrich the learning experience of North Korea's advanced students. By expanding the lexical horizon and including more complex syntactic structures, the educational authorities can offer a more comprehensive and authentic language input, potentially improving students' English proficiency more meaningfully.

There are some caveats in understanding our findings. In our research into a comprehensive analysis that covers a broad range of linguistic features, our decision to focus on three lexical features (STTR, mean word length, lexical coverage) and one syntactic feature (mean sentence length) was purposeful, as these are fundamental aspects of linguistic analyses. While we understand that these lexical features are interrelated, we chose to highlight these particular features due to their pertinence in text complexity and readability analyses. The aim of this paper was not to depict an exhaustive linguistic landscape, but rather to use these features as markers to provide a snapshot of the linguistic characteristics of North Korean English textbooks. We believe that the raw numerical values themselves are not the end of our analysis but are rather a starting point for discussing and interpreting the linguistic traits of the texts under study, thereby serving as critical markers that contribute to our understanding of the linguistic characteristics of North Korean English textbooks.

Going one step further, despite the title "English education," our study mainly investigates English textbooks used in North Korea. We believe these textbooks offer crucial insights into the larger context of English education in the region. The exploration of lexical and syntactic usage in these textbooks reflects the educational objectives and language teaching methodologies employed. Therefore, by comparing these aspects between North and South Korean English textbooks, we intended to be in effect shedding light on differing approaches to English education in both regions.

Overall, this study contributes to the growing body of research on English language education in North Korea and offers a unique perspective on the challenges and opportunities it presents. Future research could build on these findings to develop targeted pedagogical strategies and materials to help overcome the identified shortcomings, thereby fostering a richer and more effective English learning environment in North Korea's elite high schools.

References

Anthony, L. 2021. *AntConc version 3.5.8* [computer software]. Tokyo, Japan: Waseda University. Available online at https://www.laurenceanthony.net/software

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.

Davies, M. 2008. Word frequency data from the corpus of contemporary American English. English-Corpora: COCA. Available online at https://www.english-corpora.org/coca/

Goh, G. -Y. and Y. C. Lee. 2016. A corpus-based study of translation universals in English translations of Korean newspaper texts. *Cross-Cultural Studies* 45, 109-143.

Hu, M. and P. Nation. 2000. Unknown vocabulary density and reading comprehension. *Reading in a Foreign*

Language 13(1), 403-430.

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

Kim, J.-R. and S.-Y. Hwang 2018. An inter-Korean English textbook vocabulary analysis of gifted vs. general middle schools, *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology* 8(6), 223-231.

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.

KIUE. 2019. *Understanding North Korea*. Seoul: National Institute for Unification Education.

KIUE. 2020. *Understanding North Korea*. Seoul: National Institute for Unification Education.

KIUE. 2021. *Understanding North Korea*. Seoul: National Institute for Unification Education.

KIUE. 2022. *Understanding North Korea*. Seoul: National Institute for Unification Education.

Laufer, B. 1992. How much lexis is necessary for reading comprehension? In P. J. L. Arnaud and H. Bejoing, eds., *Vocabulary and Applied Linguistics*, 126-132. London: Palgrave Macmillan.

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, 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. and T.-Y. Kim. 2022. The paradigm shift in English language teaching in North Korea: A corpus-assisted analysis. *Korea Journal of English Language and Linguistic*s 22, 279-299.

Lee, Y. C. and S. Jwa. 2023. Feature importance ranking of translationese markers in L2 writing: A corpus-based statistical analysis across disciplines. *English Teaching* 78(2), 55-81.

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., B. J. Richards, N. Chipere. and P. Durán. 2004. *Lexical diversity and language development: Quantification and assessment*. UK: Palgrave Macmillan.

Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

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

Oh, S. and T.-Y. Kim. 2020. A comparison of high school English textbooks between pre- and post-2013 Revised Curriculum in North Korea. *Modern English Education* 21(1), 43-55.

Park, E. S. and Y. K. Shin. 2016. Goals and contents of English language teaching in North Korea: Insights from high school textbooks. *Modern English Education* 17(2), 91-109.

Schmitt, N., Jiang, X. and Grabe, W. 2011. The percentage of words known in a text and reading comprehension. *The Modern Language Journal* 95(1), 26-43.

Scott, M. 2016. *Wordsmith tools 7.0*. Liverpool: Lexical Analysis Software.

Examples in: English
Applicable Languages: English
Applicable Level: Secondary