



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교육학석사학위논문

Development of Clausal and Phrasal
Complexity in English Argumentative Essays
of Korean College Students

한국 대학생들의 논증적 에세이에 나타난
절과 구 복잡성의 발달

2023년 2월

서울대학교 대학원

외국어교육과 영어전공

서 나 래

Development of Clausal and Phrasal
Complexity in English Argumentative Essays
of Korean College Students

by
Narae Seo

A Thesis Submitted to
the Department of Foreign Language Education
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in Education

At the
Graduate School of Seoul National University

February 2023

Development of Clausal and Phrasal
Complexity in English Argumentative Essays
of Korean College Students

한국 대학생들의 논증적 에세이에 나타난
절과 구 복잡성의 발달

지도교수 오 선 영

이 논문을 교육학 석사 학위논문으로 제출함
2022년 12월

서울대학교 대학원
외국어교육과 영어전공
서 나 래

서 나 래의 석사학위논문을 인준함
2023년 1월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

Development of Clausal and Phrasal
Complexity in English Argumentative Essays
of Korean College Students

APPROVED BY THESIS COMMITTEE:

YOUNGSOON SO, COMMITTEE CHAIR

KITAEK KIM

SUN-YOUNG OH

ABSTRACT

Development of Clausal and Phrasal Complexity
in English Argumentative Essays of Korean College Students

Narae Seo
English Major, Dept. of Foreign Language Education
The Graduate School of Seoul National University

Studies that explore L2 writing development identify grammatical complexity as a primary discriminator for different proficiency levels of L2 writers. In the 1990s, grammatical complexity in L2 writing was often measured by clausal complexity, but the kind of complexity that has recently received particular attention is phrasal complexity. Such a move follows the recognition that clausal complexity represents the complexity of conversation and beginning levels of writing development, whereas phrasal complexity, specifically noun phrase complexity, represents the complexity of academic writing and advanced developmental levels. Some L2 writing studies, however, have yielded conflicting results, showing that phrasal features as noun modifiers have little predictive power for writing quality. One possible reason underlying these inconsistent results might be that most studies in this area have used corpus data from learners of heterogenous L1 backgrounds with no consideration for the significant effect of L1 on the use of complexity features in

L2 writing. Thus, this study analyzed essay samples produced only by L1 Korean writers to investigate whether clausal and phrasal complexity is associated with L2 writing proficiency and, if so, what developmental patterns can be observed based on complexity features that contribute substantially to the association. A qualitative analysis of student writing was followed up to provide a detailed description of proficiency-level differences, especially with respect to lexical realizations and error types associated with specific complexity features.

The corpus used in the present study contained 234 argumentative essays written by first-year college students, including 78 low-rated essays (A1 and A1+ levels of the CEFR), 78 mid-rated essays (B1 and B1+ levels of the CEFR), and 78 high-rated essays (B2+, C1, and C2 levels of the CEFR). Drawing on Biber et al.'s (2011) developmental index, the nine clausal and eight phrasal complexity features were extracted from the tagged corpus using regular expressions to measure the frequency of each feature.

The result of a Pearson Chi-square test demonstrated a statistically significant association between the three proficiency levels and the use of clausal and phrasal complexity features. The post-hoc residual analysis revealed five complexity features with great contribution to the association: finite adverbial clause, noun complement clause, WH relative clause, prepositional phrase (of), and prepositional phrase (other). Especially noteworthy is the finding that the main source of complexity at each proficiency level agrees with its corresponding developmental

stage reported by Biber et al. (2011), and thus, developmental patterns for Korean college students are successfully explained by two parameters: (1) structural form (finite dependent clauses vs. dependent phrases) and (2) syntactic function (clause constituents vs. noun phrase constituents). Specifically, the development proceeds from (i) clausal complexity mainly via finite adverbial clauses (i.e., finite dependent clauses functioning as clause constituents); through (ii) the intermediate stage of heavy reliance on WH relative clauses (i.e., finite clause types functioning as noun phrase constituents); to finally (iii) phrasal complexity primarily via prepositional phrases (of) (i.e., phrasal structures functioning as noun phrase constituents).

Surprisingly, premodifying adjectives and nouns were found to have no significant association with L2 writing proficiency despite being noun-modifying phrasal features. The subsequent qualitative analysis of student writing, however, illustrated greater proficiency of the highly rated essays in using these features in two regards. First, the lower-rated essays drew much more heavily on adjective-noun sequences presented in writing prompts than the higher-rated essays. Second, the number of errors in the composition of noun-noun sequences noticeably decreased in the higher-rated essays. The qualitative observation concerning *that* complement clauses, on the other hand, identified the reliance on a limited set of controlling nouns and conversational styles of controlling verbs in student writing across proficiency levels.

Three main pedagogical implications are provided based on the findings: (i) the

use of empirically derived developmental stages to create detailed rating scale descriptors and provide more customized writing courses on the use of complexity features; (ii) the need for classroom instruction on common academic controlling nouns and verbs used in *that* complement clauses given the importance of academically oriented lexical realizations of grammatical structures; and (iii) the need to address recurrent errors, particularly in terms of using premodifying nouns and relative clauses.

Key Words: Grammatical complexity, Academic writing, L2 writing development, Clausal complexity, Phrasal complexity

Student Number: 2021-20090

TABLE OF CONTENTS

ABSTRACT	i
TABLE OF CONTENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER 1. INTRODUCTION	1
1.1 Background of the Study	1
1.2 Purpose of the Study.....	4
1.3 Research Questions.....	5
1.4 Organization of the Thesis.....	6
CHAPTER 2. LITERATURE REVIEW	8
2.1 Grammatical Complexity in L2 Writing	8
2.1.1 Definition of Grammatical Complexity.....	9
2.1.2 Grammatical Complexity in L2 Writing Studies.....	13
2.2 Criticism of Traditional Measures of Grammatical Complexity.....	15
2.2.1 Reductiveness and Redundancy of Length- and Subordination- based Measures.....	16
2.2.2 Inappropriateness of the T-unit Approach to the Assessment of Writing Development	21
2.3 Measures of Grammatical Complexity in L2 Writing.....	24
2.3.1 Clausal and Phrasal Complexity in Relation to L2 Writing Development.....	25
2.3.2 Studies on Clausal and Phrasal Complexity in L2 Writing	31
2.4 Variation in the Use of Grammatical Complexity Features	36
2.4.1 The Effect of L1 Background.....	37
2.4.2 The Effect of Genre	43
2.4.3 The Effect of Timing Condition	46
CHAPTER 3. METHODOLOGY	50
3.1 Learner Corpus	50
3.1.1 Description of YELC 2011	50
3.1.2 Description of a Subset of YELC 2011 used in the Study.....	53

3.2 Grammatical Complexity Measures	55
3.3 Corpus Tagging and Automatic Extraction	59
3.4 Data Analysis.....	65
CHAPTER 4. RESULTS AND DISCUSSION	70
4.1 Descriptive Statistics	70
4.2 The Association between L2 Writing Proficiency and Grammatical Complexity	76
4.3 The Developmental Patterns of Grammatical Complexity	77
4.4 The Grammatical Complexity Features with Great Contribution to the Association.....	84
4.4.1 Finite Adverbial Clauses.....	84
4.4.2 Prepositional Phrases as Nominal Postmodifiers	92
4.4.3 WH Relative Clauses.....	100
4.4.4 Finite Complement Clauses Controlled by Nouns	106
4.5 The Grammatical Complexity Features with Little Contribution to the Association.....	112
4.5.1 Premodifying Adjectives	113
4.5.2 Nouns as Nominal Premodifiers.....	120
4.5.3 Finite Complement Clauses Controlled by Verbs or Adjectives	125
CHAPTER 5. CONCLUSION.....	136
5.1 Major Findings	136
5.2 Pedagogical Implications	139
5.3 Limitations and Prospect for Future Research	142
REFERENCES.....	145
APPENDICES.....	161
ABSTRACT IN KOREAN.....	165

LIST OF TABLES

Table 2.1 Complexity Dimensions used across 27 Studies on College-level L2 Writing	17
Table 2.2 Hypothesized Developmental Stages for Complexity Features.....	26
Table 3.1 Statistical Information of YELC 2011	51
Table 3.2 Description of the Part 2 Dataset in YELC 2011	52
Table 3.3 Description of the Selected Data.....	54
Table 3.4 Clause- and Phrase-level Complexity Measures used in the Analysis.	56
Table 3.5 Interpretation of Cramer's V	67
Table 4.1 Frequencies of Grammatical Complexity Features.....	71
Table 4.2 Results of the Pearson Chi-square Test and Cramer's V.....	76
Table 4.3 Results of the Adjusted Standardized Residuals	78
Table 4.4 The Observed Developmental Stages for Korean College Students....	81
Table 4.5 Top 6 Nouns Controlling <i>That</i> -clauses	109
Table 4.6 Top 6 Verbs Controlling <i>That</i> -clauses	126

LIST OF FIGURES

Figure 2.1 A Taxonomy of Complexity Constructs	10
Figure 2.2 A Multi-dimensional Representation of Syntactic Complexity.....	19
Figure 2.3 The Process of Nominalization.....	30
Figure 4.1 Frequencies of Clause-level Complexity Features	72
Figure 4.2 Frequencies of Phrase-level Complexity Features	73
Figure 4.3 Frequencies of Finite Adverbial Clauses	88
Figure 4.4 Frequencies of Premodifying Adjectives.....	115

CHAPTER 1. INTRODUCTION

This study aims to detect the developmental patterns of clausal and phrasal complexity in argumentative essays produced by Korean college students. This introductory chapter outlines the theoretical framework with the necessity for this study. First, the background of the study is presented in Section 1.1, followed by the purpose of this research in Section 1.2. Then, Section 1.3 states the research questions for the study, and lastly, Section 1.4 describes the organization of the thesis.

1.1 Background of the Study

Complexity, accuracy, and fluency (CAF) are three key dimensions characterizing qualities of language performance in second language acquisition (SLA) (Housen & Kuiken, 2009; Norris & Ortega, 2009; Wolfe-Quintero et al., 1998). Since Skehan (1989) completed the CAF triad by adding the concept of complexity, complexity has been heavily researched in second language (L2) studies as a dependent variable. For example, complexity of L2 performance has been measured in relation to L2 proficiency (e.g., Lu, 2011; Ortega, 2003; Yoon, 2017), the effects of writing topic (e.g., Yang et al., 2015; Yoon, 2017), the effects of corrective feedback (e.g., Fazilatfar et al., 2014; Sheppard, 1992), or the effects of task complexity (e.g., Ong & Zhang, 2010; Rahimpour, 2007; Révész et al.,

2017; Ruiz-Funes, 2014). Of several subdomains of L2 complexity, grammatical complexity in particular has been widely acknowledged as a reliable indicator of L2 writing development (Rimmer, 2006; Wolfe-Quintero et al., 1998). Studies that explored L2 writing proficiency in the 1990s mostly adopted clausal subordination measures and T-unit-based measures¹ as useful indices of grammatical complexity, assuming that more subordinations and extended units represent a higher degree of complexity (Ortega, 2003; Wolfe-Quintero et al., 1998). However, the increasing amount of research questioned the validity of the T-unit and clausal subordination as markers of L2 writing proficiency, with a growing consensus that noun-modifying phrasal features, rather than clausal features, better reflect the complexity of formal academic writing where extensive information should be carried in a highly compact structure (Halliday, 1989; Rimmer, 2006).

Particularly noteworthy is Biber et al.'s (2011) large-scale corpus study that provided robust empirical evidence that the complexities of conversation and academic writing are essentially different. Put simply, the former is characterized by clausal modification, whereas the latter is by phrasal modification embedded in a noun phrase (NP). Based on these findings, Biber et al. (2011) proposed the developmental stages of grammatical complexity assuming that the grammar of speech is easier and acquired earlier than that of writing.

In reply to Biber et al. (2011), a large body of L2 writing research has been

¹ T-unit indicates “a main clause with all subordinate clauses attached to it” (Hunt, 1965, p. 20).

carried out to determine whether L2 English writers also progress from clausal complexity to phrasal complexity as with first language (L1) English writers (e.g., Kim, 2020; Lan, Lucas, & Sun, 2019; Taguchi et al., 2013; Park, 2017; Parkinson & Musgrave, 2014). Most studies yielded findings in support of Biber et al.'s (2011) developmental index of complexity features, showing that noun-modifying phrasal features are highly correlated with L2 writing proficiency, but the results of some studies seemed incongruent with Biber et al. (2011), suggesting that NP-related features are not reliable predictors of proficiency-level differences (Crossley & McNamara, 2014; Guo et al., 2013; Park, 2017; Yang et al., 2015). The inconsistency might be attributed to the research design that failed to consider the variables affecting the relationship between the use of complexity features and writing proficiency levels. Indeed, the overwhelming majority of the studies in this area analyzed data from learners with heterogeneous L1 backgrounds, although L1 influence on the use of complexity features has been empirically proven (Crossley & McNamara, 2012; Kang, 2005; Lu & Ai, 2015; Staples & Reppen, 2016). The main point of previous research on the effects of L1 backgrounds on grammatical complexity in L2 writing is that L2 proficiency alone cannot successfully predict the complexity preferred in L2 writing, and vice versa. This is because the characteristics of learners' L1 are likely to bring preference or avoidance for specific L2 features, and thus, learners with specific L1s might have different developmental patterns and may not conform to proficiency-based

prediction (see more detailed discussion in Section 2.4.1).

Given the potential impact of L1 on grammatical complexity, it is no surprise that studies of L2 writing development based on data from learners of various L1 backgrounds showed disagreement on specific complexity features. Further, it would be unreasonable to generalize any of these findings to a specific L1 group, and in this sense, more research based on a single learner population is warranted. No studies that the researcher knows of have compared clause-level and phrase-level complexity features produced by L1 Korean learners in relation to L2 writing proficiency. Thus, this paper attempted to investigate the language produced by Korean college students at the clausal and phrasal level drawing on Biber et al.'s (2011) developmental indices to elucidate their developmental patterns of grammatical complexity.

1.2 Purpose of the Study

The main goal of this research is threefold: to assess the extent to which clause-level and phrase-level complexity features are associated with the quality of argumentative essays of Korean college students; to identify developmental patterns based on complexity features that contribute greatly to the association; to clarify the differences among proficiency groups in the use of complexity features by qualitatively analyzing student writing. In doing so, this study ultimately aims at estimating the developmental trajectory of grammatical complexity for Korean

college students with reference to the developmental stages hypothesized by Biber et al. (2011).

Understanding the developmental stages of Korean learners may inform writing instruction by suggesting crucial areas that require pedagogical focus to move on to the next stage depending on learners' writing proficiency. Contributing to writing instruction is realized in two specific ways. One important contribution lies in useful quantitative information on the kinds of complexity features generally favored by learners at specific proficiency levels and how they differ from those typically associated with advanced academic writing. Another valuable contribution is made by the qualitative analysis that illustrates differences among proficiency levels in the lexical realizations of complexity features and how far their lexical choices deviate from academic norms. The qualitative check also helps reveal the error types that learners at a particular proficiency level often make. Efforts can be made to give fine-tuned writing instruction based on such quantitative and qualitative information with a view to enriching students' repertoire of academically oriented grammatical structures and associated lexical items and improving accuracy in using these lexico-grammatical patterns.

1.3 Research Questions

There are three primary purposes that this thesis seeks to attain by examining complexity features in argumentative essays of Korean college students at three

differing proficiency levels. The first is to investigate the extent to which the occurrence rates of clausal and phrasal features are associated with L2 writing proficiency of Korean college students. The second is to identify developmental patterns based on complexity features found to be the primary discriminators of proficiency-level differences. The third is to reveal qualitative differences among proficiency levels especially with regard to lexical choices and error types associated with specific complexity features. In accordance with these purposes, research questions are formulated as follows:

1. Is there an association between L2 writing proficiency and clausal and phrasal complexity in English argumentative essays written by Korean college students?
2. What developmental patterns can be observed based on complexity features that contribute to the association the most?
3. What differences in the actual use of complexity features exist across proficiency levels?

1.4 Organization of the Thesis

This thesis is composed of five chapters. Following this introductory chapter, the general theoretical background is surveyed in Chapter 2. It first introduces the key concept of this study, i.e., grammatical complexity, and a critical evaluation of traditional complexity measures followed by their proper alternatives. This

chapter further elaborates on several factors that influence the values of complexity measures. Chapter 3 explains the research method, including the description of the data used in this study, measures of grammatical complexity used to analyze the data, the process of extracting target complexity features, and data analysis methods adopted in this study. Chapter 4 reports the findings of this study by answering the three research questions. The quantitative findings on specific complexity features are further examined by qualitatively analyzing student writing samples. Finally, Chapter 5 summarizes the major findings of this research followed by the presentation of pedagogical implications and limitations of this research that offer suggestions for future studies.

CHAPTER 2. LITERATURE REVIEW

This chapter introduces the notion of grammatical complexity and its relationship to L2 writing development in Section 2.1. Then, two central criticisms leveled against traditional measures of complexity are discussed in Section 2.2, and alternative measures better suited for L2 writing development are presented in Section 2.3. Lastly, several factors that affect the use of grammatical complexity features are illustrated in Section 2.4.

2.1 Grammatical Complexity in L2 Writing

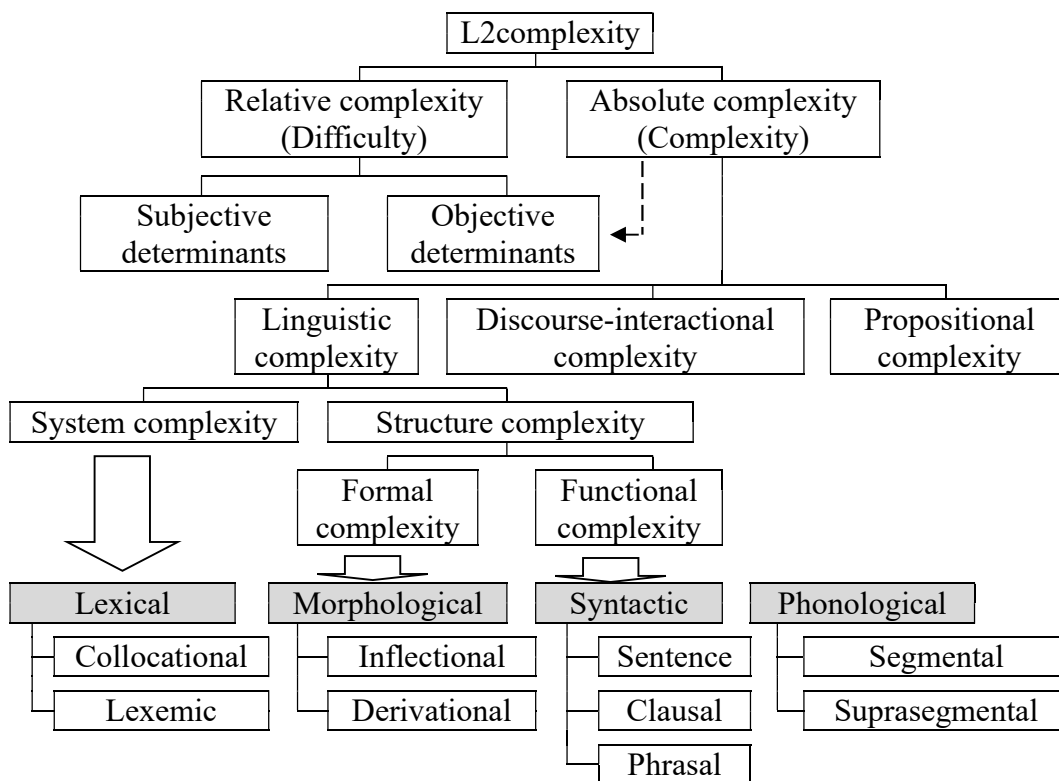
Grammar constitutes a principal factor in gauging writing quality and L2 writers' global proficiency (Cumming & Mellow, 1996; McNamara, 1990). It consists of two aspects that can be measured: accuracy and range, or, more specifically, "control of structures and freedom from error" and "the variety of grammatical structures that test-takers employ" (Rimmer, 2006, p. 498). Of particular relevance to the current discussion of grammatical complexity is the range. The definition of grammatical complexity is specified in Section 2.1.1 based on Bulté and Housen's (2012) taxonomic model of L2 complexity, and its multidimensional nature (i.e., theoretical, observational, operational) is unraveled. Then, a brief survey of previous research on how grammatical complexity is associated with L2 writing proficiency is provided in Section 2.1.2.

2.1.1 Definition of Grammatical Complexity

As noted in Section 1.1, CAF has long been viewed as key traits that are largely reflective of L2 proficiency, and much research has been done on measures of CAF development (Bulté & Housen, 2012; Housen & Kuiken, 2009; Norris & Ortega, 2009; Wolfe-Quintero et al., 1998). The major imponderable and vague dimension of the CAF triad is complexity because it can be interpreted in several distinct ways (Housen & Kuiken, 2009). In the first place, the term can be used to describe either “properties of language task (task complexity)” or “properties of L2 performance and proficiency (L2 complexity)” (Housen & Kuiken, 2009, p. 463). L2 complexity, which is of interest to this research, is further subdivided into two components: relative complexity (or simply difficulty) and absolute complexity (or simply complexity) as represented in Figure 2.1 (Bulté & Housen, 2012). The relative complexity is “a variable property” with the two determining factors, namely, subjective learner-internal determinants (e.g., aptitude, motivation, memory capacity, L1 background) and objective learner-external determinants (e.g., the saliency of an L2 feature in language input, its frequency, its inherent complexity) (Housen & Kuiken, 2009, p. 463). As the dotted line in Figure 2.1 indicates, a portion of objective determinants is accounted for by absolute complexity, i.e., language complexity quantified by the number of individual components that make up an L2 feature and of their connections (Bulté & Housen, 2012). Thus, technically, the notion of relative complexity embraces

absolute complexity as one of many determinants that cause learning difficulties.

FIGURE 2.1 A Taxonomy of Complexity Constructs



Note. From Bulté and Housen (2012, p.23).

As further illustrated in Figure 2.1, absolute complexity, in turn, is classified into three factors: linguistic, discourse-interactional, and propositional. Of relevance to this study is linguistic complexity, which has two different interpretations: “a dynamic property of the learner’s L2 system” (i.e., system complexity) and “a more stable property of the individual linguistic items, structures or rules that make up the learner’s L2 system” (i.e., structure complexity)

(Bulté & Housen, 2012, p. 25). System complexity involves “the degree of elaboration, the size, breadth, width, or richness of the learner’s L2 system,” whereas structure complexity, consisting of functional and formal complexity, is more relevant to “depth” rather than to “breadth or range” (Bulté & Housen, 2012, p. 25; Housen & Kuiken, 2009).

When looking into linguistic complexity, its diverse subareas (i.e., lexical, morphological, syntactic, phonological) are subject to the assessment, as indicated in Figure 2.1. Among the four subconstructs, the area of interest in this research is syntactic complexity,² more specifically, phrasal and clausal complexity. The subconstructs of linguistic complexity could be studied at three distinct dimensions: theoretical, observational, and operational (Bulté & Housen, 2012). Bulté and Housen (2012) characterized (1) the theoretical dimension as an abstract property of a structure; (2) the observational dimension as its manifestation in language performance on various levels; and (3) the operational dimension as the quantifiable measures designed to provide an objective value for the degree of complexity of a specific writing sample. A comparable scheme was designed by Norris and Ortega (2003) for the assessment process, which begins from “construct definition” through “behavior identification” and finally leads to “observation scoring” (p. 720). In their framework, measurement is argued to

² Bulté and Housen (2012) laid down an expansive definition of grammatical complexity by incorporating both morphological and syntactic complexity as opposed to many other L2 writing studies where the term grammatical complexity is used interchangeably with the term syntactic complexity (e.g., Biber et al., 2011; Lu, 2011).

serve as a bridge between observable phenomena and theoretical interpretations of those phenomena. Given this intimate connection, the three dimensions of complexity should be made explicit to evaluate language complexity in a valid and reasonable way. Indeed, Wolfe-Quintero et al. (1998), Norris and Ortega (2009), and Biber et al. (2011) pointed out a significant mismatch between what researchers of L2 writing development want to assess and the complexity measures they employ. It follows then that for this study, there must first be a theoretical definition of grammatical complexity. It should then be established in what linguistic categories (e.g., T-units, subordinations, clauses, phrases) it can be observed. Based on these two preliminary levels of definitions, finally, the quantitative measures to operate grammatical complexity (e.g., length, ratio, frequency) have to be set up to provide objective numerical values indicative of the degree of complexity.

Accordingly, the following section undertakes a brief review of previous studies that present the theoretical definitions of grammatical complexity in relation to writing proficiency. Then, from the observational perspective, the shift of the primary focus from T-units and subordinations to phrasal features associated with complex NPs is discussed in Section 2.2. Building on the preceding discussions, the operational dimension concerning how to measure grammatical complexity in L2 written production is examined in detail in Section 2.3.

2.1.2 Grammatical Complexity in L2 Writing Studies

There has been extensive research on L2 writing development that adopted grammatical complexity as a strong development indicator (Ortega, 2003). Accordingly, most researchers in this field have doubtlessly employed measures of grammatical complexity to assess writing proficiency. Nonetheless, the notion of grammatical complexity remains rather vague in L2 writing studies (Bulté & Housen, 2012; Lan, Liu, & Staples, 2019; Rimmer, 2006). Considering that “a definition of grammatical complexity can be based on the usual understandings of a complex system as one consisting of many different elements each with a number of degrees of freedom” (Nichols, 2009, p. 111), it comes as no surprise that studies of grammatical complexity failed to establish its unified and transparent definition. For instance, Wolfe-Quintero et al. (1998) defined grammatical complexity as “a wide variety of both basic and sophisticated structures” (p. 69). However, Lan, Liu, and Staples (2019) cast doubt on whether basic structures can be an index of complexity. Besides, clarification is required as to what makes structures sophisticated. Rimmer (2006) listed several factors that are responsible for grammatical complexity, including “sentence length, embedding, ellipsis, markedness, and register,” the accumulative effects of which can result in strikingly increased complexity. This comprehensive definition was based on the awareness of the problem with traditional complexity measures that fail to capture a wide variety of grammatical structures, but problems still remain in respect of

the need for elaboration on each factor.

The definitions of other researchers also pose similar problems in their elusiveness. They, however, describe grammatical complexity of L2 writing from two common perspectives, which are parallel to the two constructs of grammatical complexity introduced by Bulté and Housen (2012): system complexity (i.e., grammatical diversity) and structure complexity (i.e., grammatical sophistication). For instance, Ortega (2003) conceptualized grammatical complexity as “the range of forms that surface in language production and the degree of sophistication of such forms” (p. 492). Rimmer (2006) conveyed a similar perspective when introducing the term range as one of the measurable dimensions of grammar. Specifically, range was argued to be determined by both quantitative and qualitative components: “the number of different grammatical forms used and their complexity” (p. 498). Lu (2011) also conceived grammatical complexity as consisting of two facets: syntactic variation and sophistication. The former indicates “the range of syntactic structures that are produced,” whereas the latter indicates “the degree of sophistication of such structures” (p. 36). Following the notion of grammatical complexity suggested by Ortega (2003) and Lu (2011), Crossley and McNamara (2014) associated grammatical complexity with “the sophistication of syntactic forms produced by a speaker or writer and the range or variety of syntactic forms produced” (p. 67).

The common assumption underlying the definitions of grammatical

complexity in L2 writing studies is that L2 writers use increasingly more diverse and sophisticated grammatical structures as they advance in language proficiency. In other words, the diversity of grammatical structures and their degree of sophistication can be useful indicators of writing proficiency differences. On the basis of these theoretical assumptions, the following section discusses the appropriate linguistic units in which grammatical diversity and sophistication of L2 writing can be best captured.

2.2 Criticism of Traditional Measures of Grammatical Complexity

The vast majority of the early developmental studies adopted measures based on clausal subordination or T-unit to evaluate grammatical complexity, on the premise that more advanced writers produce increasing numbers of subordination and extended units (e.g., Becker, 2010; Brown et al., 2005; Ellis & Yuan, 2004; Grant & Ginther, 2000; Larsen-Freeman, 2006; Li, 2000; Nelson & Van, 2007; Norrby & Håkansson, 2007). In fact, Wolfe-Quintero et al. (1998), in their research synthesis on L2 writing development of the late 1990s, found mean number of clauses per T-unit (C/TU) and mean number of dependent clauses per clause (DC/C) to be the best indices of grammatical complexity in L2 writing.

However, a growing number of recent research has criticized these traditional measures, raising the validity issue of whether they reliably reflect L2 writing

proficiency. Two central criticisms have been leveled against these complexity measures: first, the reductive and redundant nature of length- and subordination-based measures, and second, the inappropriateness of T-unit analysis as a measure of grammatical complexity in written discourse. They are discussed in turn in Sections 2.2.1 and 2.2.2.

2.2.1 Reductiveness and Redundancy of Length- and Subordination-based Measures

The widespread use of holistic measures based on length or amount of subordination is well documented by Ortega (2003), who conducted a meta-analysis of prior L2 writing research on syntactic complexity. Of the total 27 studies examined in her research synthesis, 25 employed mean length of T-unit (MLTU), either as a single indicator of writing proficiency or in conjunction with mean length of clause (MLC), mean length of sentence (MLS), mean number of clauses per T-unit (C/TU), mean number of T-units per sentence (TU/S), mean number of dependent clauses per clause (DC/C), and mean number of dependent clauses per T-unit (DC/TU). As shown in Table 2.1, these complexity measures can be categorized into four columns according to the associated complexity dimension. The overall complexity measures are listed in Column 1, and the measures of complexity by coordination in Column 2, by subordination in Column 3, and by phrasal elaboration in Column 4.

TABLE 2.1
Complexity Dimensions used across 27 Studies on College-level L2 Writing

Study	Overall complexity	Complexity by coordination	Complexity by subordination	Phrasal complexity
Cross-sectional				
Bardovi-Harlig and Bofman (1989)			C/TU	
Flahive and Snow (1980)	MLTU		C/TU	
Gaies (1976)	MLTU		C/TU	MLC
Homburg (1984)	MLTU, MLS	TU/S	C/TU	
Ho-Peng (1983)	MLTU			
Kameen (1979)	MLTU, MLS		C/TU, DC/C	MLC
Larsen-Freeman (1978)	MLTU			
Larsen-Freeman (1983-Study 2)	MLTU			
Larsen-Freeman and Strom (1977)	MLTU			
Perkins (1980)	MLTU		C/TU	
Perkins and Homburg (1980)	MLTU			
Sharma (1980)	MLTU		C/TU	MLC
Tedick (1990)	MLTU			
Hirano (1991)			C/TU, DC/TU	
Neff et al. (1998)	MLTU		C/TU	MLC
Nihalani (1981)	MLTU			
Cooper (1976)	MLTU, MLS	TU/S	C/TU	MLC
Dvorak (1987)	MLTU			
Henry (1996)	MLTU			
Kern and Schultz (1992)	MLTU			
Monroe (1975)	MLTU, MLS	TU/S	C/TU	MLC
Longitudinal				
Arthur (1979)	MLTU			
Larsen-Freeman (1983-Study 3)	MLTU			
Arnaud (1992)	MLTU			
Casanave (1994)	MLTU			
Ishikawa (1995)	MLTU			
Kern and Schultz (1992)	MLTU			

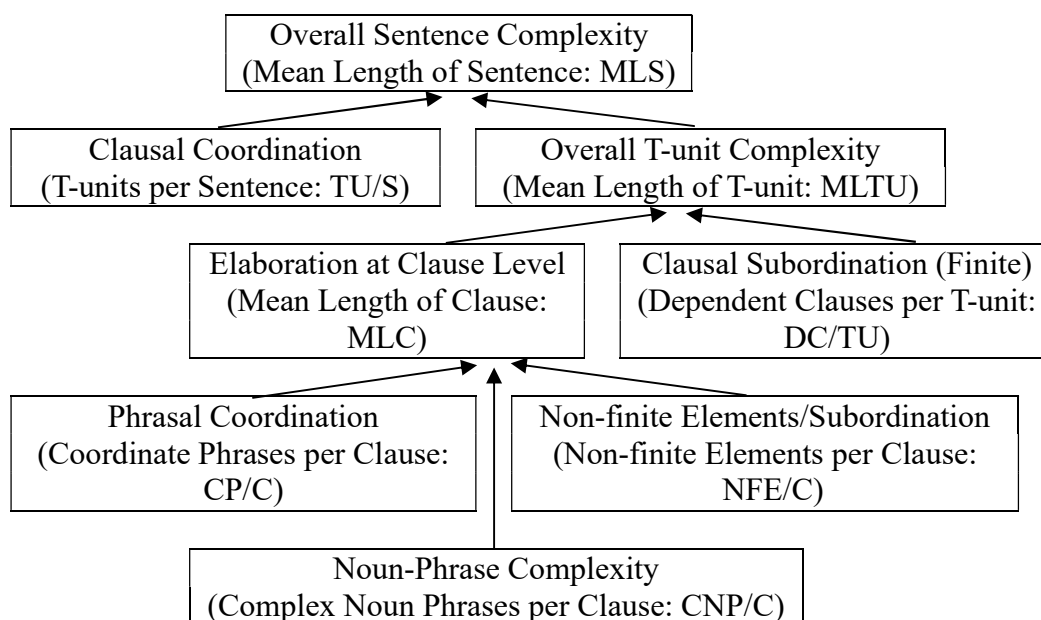
Note. Adapted from Ortega (2003, p. 497) and Norris and Ortega (2009).

The measures organized in Table 2.1 appear to effectively capture different sources of complexity, but these traditional indices have two major problems: reductiveness and redundancy. Firstly, length-based measures like MLTU and MLS in Column 1 are not sensitive to internal structures that contribute to the length of T-units or sentences by providing only a few large-grained variables (Biber et al., 2011; Kyle & Crossley, 2018; Norris & Ortega, 2009; Rimmer, 2006; Wolfe-Quintero et al., 1998). In other words, MLTU scores merely provide “a general indication of how elaborated a particular main clause is but say nothing about the type(s) of elaboration included” (Kyle & Crossley, 2018, p. 102). Thus, Biber et al. (2020) viewed these length-based calculations as “omnibus measures” that combine several distinct structural, syntactic, and functional characteristics into a single quantitative estimate (p. 2).

The reductive nature of MLS and MLTU is well displayed in Figure 2.2, which visually represents the hierarchical structures among subdimensions of syntactic complexity. In this figure, MLS and MLTU are located at higher levels as indices of the overall sentence and T-unit complexity, respectively, “essentially encapsulating all sub-constructs” (e.g., MLC, DC/TU, CNP/C) (Yang et al., 2015, p. 55). As global complexity measures that fail to clarify specific grammatical features involved in elaboration, MLS or MLTU scores are not linguistically informative and interpretable. Such holistic length-based measures only serve to demonstrate overall linguistic complexity. In this sense, it may be reasonable to

suppose that any indices involving mean number of production unit are measures of fluency rather than complexity (Norris & Ortega, 2009; Wolfe-Quintero et al., 1998). Indeed, Ishikawa (2007) and Storch and Wigglesworth (2007) employed MLTU in their L2 writing studies as an index of fluency, not complexity.

FIGURE 2.2 A Multi-dimensional Representation of Syntactic Complexity



Note. From Yang et al. (2015, p.54).

The measure of MLC in Column 4 of Table 2.1 is clearly distinguished from other length-based measures for its numerical result is independent of the amount of subordination and determined solely by subclausal complexification such as phrasal modification and nominalization. MLC scores, however, are still open to interpretation because clauses can be lengthened in several different ways. In other words, MLC scores give no information on whether increased phrasal complexity

is due to the addition of complements or phrasal embedding (Kyle & Crossley, 2018; Norris & Ortega, 2009).

Second, the indices which tap complexity via amount of subordination in Column 3 of Table 2.1 (i.e., C/TU, DC/C, DC/TU) are redundant in that they measure precisely the same kind of complexity in the end (Lu & Ai, 2015; Norris & Ortega, 2009). The reason is that all these subordination measures have clauses or dependent clauses as the numerator in common, and the sole difference lies in the selection of denominator. Thus, “the only way in which values for this family of measures would show an increase is when more subordinate or dependent clauses are produced” (Norris & Ortega, 2009, p. 560). Accordingly, one measure of the subordination family is no different from the replication of another subordinate measure except for at what level the ratio is computed. Thus, employing more than one in this family of measures would cause redundancy and multicollinearity problems. The measure of clausal coordination in Column 2 (i.e., TU/S) was introduced as an alternative to subordination measures for beginning levels of development since coordination measures are thought to better capture the source of complexity produced by low-proficiency L2 writers (Bardovi-Harlig, 1992; Norris & Ortega, 2009). That being so, measurement via the amount of coordination provides a very limited view of complexity and cannot have a predictive value for L2 writers of higher proficiency. For that reason, coordination measures were rarely used compared to other measures, as seen in Table 2.1.

To summarize this section, there is redundancy in clause-based measures used in previous L2 writing studies in that they tap the identical complexity dimension. More important, the reductive nature of length-based measures fails to provide a clear view of clausal and phrasal complexity by oversimplifying the complexity involved in written language production. Thus, alternative measures are required to adequately represent the diverse sources of complexification and to distinguish performances at different proficiency levels.

2.2.2 Inappropriateness of the T-unit Approach to the Assessment of Writing Development

Besides reductiveness and redundancy, the other criticism toward traditional measures of complexity is that the T-unit approach is not linguistically well-grounded by confounding various linguistic features with distinct distributions and functions (Biber et al., 2011; Biber et al., 2020). There has been a wealth of research arguing that clausal subordination is a typical characteristic of spoken registers, whereas complex NPs are essential characteristics of written academic registers (Halliday, 1989, 2004; Rimmer, 2006). For instance, Halliday (1989) argued that nouns and nominalization are the hallmarks of advanced writing, saying that “the highly information-packed, lexically dense passages of writing often tend to be extremely simple in their grammatical structure” (p. 87). Similarly, Rimmer (2006) pointed out that a short sentence can be “grammatically denser, in

the sense that functional words are omitted, in order to pack a lot of information into a short compass” (p. 506).

Consistent with such a theoretical foundation, corpus-based research on register variation confirmed that T-unit-based measures capture the complexities typical of informal conversation and that the complexities of advanced academic writing are fundamentally different (Biber, 1985, 1986, 1992; Biber et al., 1999; Biber, Conrad, Reppen, et al., 2002; Biber et al., 2011). Thus, evaluating complexity based on a single uniform standard without regard to register variation and functional resources associated with grammatical complexity could lead to the misinterpretation, for example, that Sentence (1) is more complex than Sentence (2) (see below). According to the corpus findings, a reasonable interpretation is to assume that the two sentences are complex in completely different ways.

(1) Well, since he got so upset, I just didn't **think** we would want to wait for Tina to come back. [from a conversation]

- T-unit length: 20
- Number of dependent clauses per T-unit: 4

(2) This may **be** part of the reason for the statistical link between schizophrenia and membership in the lower socioeconomic classes. [from a university textbook]

- T-unit length: 20
- Number of dependent clauses per T-unit: 0

Note. Adapted from Biber et al. (2011, p. 14).

The main verbs are in bold and underlined, and the verbs of dependent clauses are underlined.

The main point of Biber et al.'s (2011) corpus study, which reported the

grammatical features typical of academic writing when compared with informal conversation, was that “the complexity of conversation is clausal, whereas the complexity of academic writing is phrasal” (p. 22). More specifically, two major parameters of register variation were identified in their study: (1) structural type (finite dependent clauses vs. dependent phrases) and (2) syntactic function (clause constituents vs. NP constituents) (see below). Particularly noteworthy was that the two parameters were combined in the opposite direction in the interpersonal spoken register and formal written register. That is, the most common complexity features in conversation are finite dependent clauses that syntactically function as clause constituents, while the most favored types in academic writing are phrasal structures that syntactically function as NP constituents.

Favored in conversation	↔	Favored in academic writing
Parameter A: Structural type		
finite dependent clauses	vs.	dependent phrases (nonclausal)
Parameter B: Syntactic function		
constituents in clauses	vs.	constituents in noun phrases

Note. From Biber et al. (2011, p. 22).

The presence of an oral-literate dimension identified via Multi-Dimensional (MD) analysis provided a further rationale for the significance of phrasal complexity as a measure of writing proficiency (Biber, 2006; Biber et al., 2016). The MD analysis explored the dimensions of linguistic variation assuming that “linguistic co-occurrence patterns have a functional basis” (Biber et al., 2016, p.

658), as many scholars have argued (e.g., Halliday, 1989; Nichols, 1984; Rimmer, 2006). The oral-literate dimension detected in the MD analysis consists of positive and negative features in complementary distribution, confirming the finding of Biber et al. (2011) that most grammatical characteristics are “strongly favored in either conversation or academic writing, but not both” (p. 22). The oral pole and the literate pole of this dimension are respectively associated with clausal and phrasal complexity features.

As such, the cumulative empirical findings clearly indicate that what reliably mirrors grammatical complexity in academic written register is phrasal complexity rather than clausal complexity. Given the influence of registers on language use, the application of clausal complexity indices based on subordination or T-unit to gauge the complexity of written language production is seriously misguided.

2.3 Measures of Grammatical Complexity in L2 Writing

As discussed in the preceding section, traditional measures of complexity via subordination or T-unit fail to tap the dimension of complexity typical of written academic texts. Thus, alternative measures to evaluate writing development need to be established. Section 2.3.1 presents the complexity measures appropriate for the study of L2 writing development, drawing on Biber et al.’s (2011) corpus-based register research in tandem with the theoretical justification for their argument. Then, Section 2.3.2 introduces a body of empirical research performed

to validate the developmental stages for complexity features proposed by Biber et al. (2011).

2.3.1 Clausal and Phrasal Complexity in Relation to L2

Writing Development

On the assumption that grammatical constructions frequently used in conversation are acquired readily and earlier than those commonly used in academic writing, Biber et al. (2011) reasoned that the grammatical structures typical of formal written texts represent higher levels of production complexity compared to those typical of informal conversation. This interpretation led to the proposal for five stages of complexity development, as presented in Table 2.2. These five developmental stages can be boiled down to three stages by adopting the aforementioned two parameters of register variation (i.e., structural type, syntactic function) as determinants for progression. Therefore, the developmental stages can be summarized as below:

“Thus, the stages generally progress from finite clauses functioning as constituents in other clauses, through intermediate stages of nonfinite dependent clauses and phrases functioning as constituents in other clauses, and finally to the last stage requiring dense use of phrasal (nonclausal) dependent structures that function as constituents in noun phrases” (Biber et al., 2011, pp. 29-30).

TABLE 2.2
Hypothesized Developmental Stages for Complexity Features

Stage	Grammatical structures	Examples
1	Finite complement clauses controlled by extremely common verbs	Just think <u>that he didn't pay attention.</u>
2	Finite complement clauses controlled by a wider set of verbs Finite adverbial clauses Nonfinite complement clauses controlled by common verbs Adverbs as adverbials Attributive adjectives as nominal premodifiers	I'd forgotten <u>that he had just testified on that one.</u> I'm assuming I gained weight <u>because things are a little tighter than they used to be.</u> I don't want <u>to fight with them about it.</u> He's so confused <u>anyway.</u> It certainly has a <u>nice</u> flavor.
3	Prepositional phrases as adverbials Finite complement clauses controlled by adjectives Nonfinite complement clauses controlled by a wider set of verbs <i>That</i> relative clauses Nouns as nominal premodifiers Possessive nouns as nominal postmodifiers <i>Of</i> phrases as nominal postmodifiers Prepositional phrases other than <i>of</i> as nominal postmodifiers (concrete/locative)	He seems to have been hit <u>on the head.</u> I was sure <u>that I could smooth over our little misunderstanding.</u> The snow began <u>to fall again.</u> The guy <u>that made that call</u> Some really obscure <u>cable</u> channel <u>Tobie's</u> voice Editor <u>of the food section</u> House <u>in the suburbs</u>
4	Nonfinite complement clauses controlled by adjectives Extraposited complement clauses Nonfinite relative clauses	These will not be easy <u>to obtain.</u> It is clear <u>that much remains to be learned.</u> The method <u>used here</u> should suffice.
5	Prepositional phrases other than <i>of</i> as nominal postmodifiers (abstract) Preposition + nonfinite complement clause Complement clauses controlled by nouns Appositive noun phrases Multiple prepositional phrases as nominal postmodifiers	The specific growth rate <u>at small population sizes</u> The idea of <u>using a Monte Carlo approach</u> The hypothesis <u>that female body weight was more variable</u> The CTBS <u>(the fourth edition of the test)</u> was administered in 1997-1998. The [presence of layered <u>[[structures] at the [[[borderline]]] of cell territories]]]</u>

Note. Adapted from Biber et al. (2011, pp. 30-31).

Considering that grammatical complexity in L2 writing is commonly defined as twofold: grammatical diversity and sophistication (Bulté & Housen, 2012; Crossley & McNamara, 2014; Lu, 2011; Ortega, 2003; Rimmer, 2006), and that these theoretical constructs of complexity can be adequately observed at the level of phrases rather than clauses as proficiency develops (Biber, 2006; Biber et al., 2011; Biber et al., 2016; Halliday, 1989, 2004; Rimmer, 2006), Biber et al.'s (2011) index can be argued to offer concrete and objective measures to operate grammatical complexity of academic writing at different proficiency levels. In other words, multidimensional and developmental construct of grammatical complexity is well represented in Biber et al.'s (2011) framework by including both clause-level and phrase-level structures (Norris & Ortega, 2009). This enables complexity analyses based on frequencies of grammatical features of interest, which index both the diversity and the degree of sophistication of grammatical structures. In general, the grammatical structures at the lower end of the developmental scale represent clause-level complexity, which is deemed easier to acquire and less sophisticated, whereas those at the higher end of the developmental scale represent phrase-level complexity, which is considered more difficult to acquire and highly sophisticated.

As such, complexity measures in Biber et al.'s (2011) index are complementary as they capture different kinds of complexity reflective of different stages of development; accordingly, they should be employed and interpreted in

tandem to gain a fuller picture of L2 writing development. This is well supported by Norris and Ortega's (2009) work, which stressed the value of "devising measures that include a wide range of developmentally ranked structures regardless of their status as target-like or nontarget-like, so as to help researchers characterize L2 production that ranges along the full developmental continuum" (p. 567).

Biber et al.'s (2011) model is not only empirically motivated but well-grounded on a theoretical basis. Its theoretical underpinnings can be located in Systemic Functional Linguistics (SFL; Halliday & Matthiessen, 1999). The SFL theory postulates that the development of language complexity involves the shift from dynamic styles to synoptic styles (Norris & Ortega, 2009; Ortega, 2012). In other words, language learners are expected to proceed from dynamic to synoptic styles as they become more complex and sophisticated language users. Dynamic styles feature the dense use of coordination and, at the advanced levels of complexity within this style, subordination. At even higher levels of development, however, dynamic styles give way to synoptic styles, which are marked by the extensive use of nominalization and grammatical metaphor for carrying highly compressed information (Ortega, 2012). As shown in the following excerpt, Halliday and Matthiessen (2004) associated grammatical metaphor, primarily through nominalization, with developmentally advanced writers in the same manner that Biber et al. (2011) associated complex NPs with more proficient

writers based on register research.

“Unlike interpersonal metaphor, the other type of grammatical metaphor, ideational metaphor,³ is learned later by children and is not part of the grammar of ordinary, spontaneous conversation that children meet in the home and neighborhood; rather, it is associated with the discourses of education and science, bureaucracy and the law. Children are likely to meet the ideational type of metaphor when they reach the upper levels of primary school; but its full force will only appear when they begin to grapple with the specialized discourse of subject-based secondary education” (Halliday & Matthiessen, 2004, p. 636).

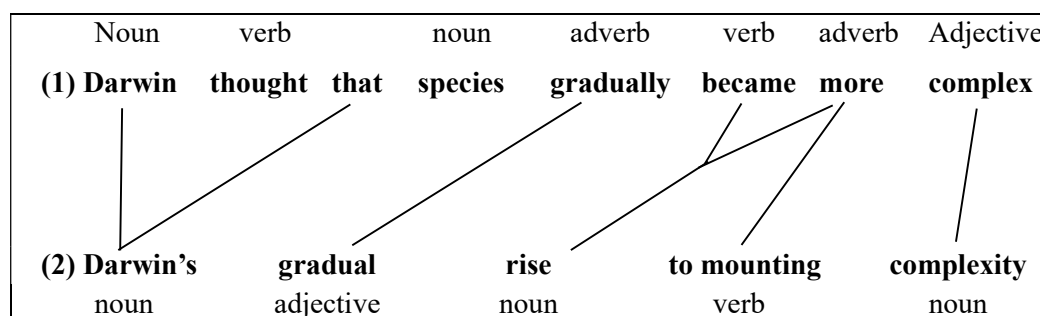
Similar to corpus-based register research (e.g., Biber, 1988, 2006), SFL theory associates the linguistic differences between dynamic and synoptic styles with different communication circumstances. The theory asserts that dynamic styles represent complexity typical of everyday conversation, whereas synoptic styles represent complexity prevalent in formal academic writing (Ortega, 2012). This claim aligns precisely with Biber et al.’s (2011) findings in their empirical research that complexity devices of spoken registers and formal written registers are entirely different, and that clausal subordination is typical of conversation, whereas academic writing is featured by complex NPs.

The two expressions with the same meaning in Figure 2.3 illustrate the advance from subordination-based complexity to nominalization-based

³ Halliday and Matthiessen (2004) categorized grammatical metaphor as either interpersonal or ideational metaphor. Interpersonal metaphor concerns mood or modality, whereas ideational metaphor is primarily associated with nominalization, which is a concern of this study.

complexity (Halliday & Martin, 1993). The eight-word semantically more explicit T-unit in (1) is the paraphrased version of the biologist Stephen Jay Gould’s grammatical metaphor regarding Darwin’s theory of evolution in (2).

FIGURE 2.3 The Process of Nominalization



Note. Adapted from Halliday and Martin (1993, p. 35).

The verb *became* and the adjective *complex* in (1) correspond respectively to the nouns *rise* and *complexity* in (2). In other words, the process (*became*) and the attribute (*complex*) are expressed by means of nouns rather than congruent grammatical categories (i.e., a verb and an adjective). Such “non-prototypical or grammatically non-congruent pairings,” by which a process, attribute, or proposition is re-constructed as “static and detached things,” are referred to as grammatical metaphor in SFL theory (Ortega, 2012, p. 144).

In sum, the language development along the dynamic-synoptic style continuum in SFL theory affords a strong theoretical basis for the developmental progression hypothesized by Biber et al. (2011). Taken together, both theoretical and empirical works yielded the consistent prediction that grammatical complexity

develops from relying on clausal subordination to phrasal elaboration, specifically in NPs.

2.3.2 Studies on Clausal and Phrasal Complexity in L2

Writing

Motivated by Biber et al.'s (2011) developmental index of complexity features, numerous studies (e.g., Kim, 2020; Kyle & Crossley, 2018; Lan, Lucas, & Sun, 2019; Taguchi et al., 2013; Park, 2017; Parkinson & Musgrave, 2014) have set out to empirically test if the developmental progression from clausal to phrasal complexity is also observed in L2 writers. One such attempt is demonstrated by Taguchi et al. (2013), who analyzed 54 placement exam essays from two proficiency levels to identify linguistic features that distinguished high-rated and low-rated essays. Complexity measures used in this study were categorized into two levels: clausal and phrasal levels. They observed that the low-rated essays were more complex than the high-rated ones with regard to the frequency counts of clausal complexity features as they showed a similar or higher frequency in five out of six clausal features analyzed. However, the exact opposite picture emerged regarding the frequency counts of phrasal complexity features. The most striking differences between the two proficiency levels were found in attributive adjectives and post-noun-modifying prepositional phrases, both of which occurred with much higher frequency in the high-rated essays. Similar findings were yielded in

Kyle and Crossley's (2018) study, which investigated 480 argumentative essays from low- and high-proficiency groups using three types of complexity measures: traditional indices, fine-grained clausal indices, and fine-grained phrasal indices. The combined predictor model demonstrated that fine-grained phrasal complexity associated with prepositional object dependents, direct object dependents, and nominal subject dependents made up a substantial amount of the variance in holistic essay scores. The findings of Taguchi et al. (2013) and Kyle and Crossley (2018) uphold Biber et al.'s (2011) work that found a close relationship between advanced writing and the use of phrasal modifiers functioning as NP constituents.

Reducing the scope of analysis, Parkinson and Musgrave (2014) looked specifically at NP complexity features using Biber et al.'s (2011) index for noun modifiers. They compared writings of English for Academic Purposes (EAP) students in preparation for graduate studies with writings of matriculated MA students already engaged in graduate studies in terms of using nominal modifiers. By comparison with the more advanced MA writing, the less proficient EAP writing showed a significantly higher proportion of attributive adjectives while demonstrating a significantly lower proportion of premodifying nouns and prepositional phrases other than *of*, in particular those with abstract meaning. These findings lend support to Biber et al.'s (2011) argument that attributive adjectives are placed in the lowest developmental stage for nominal modifiers (Stage 2), whereas premodifying nouns and postmodifying prepositional phrases

are placed in the higher developmental stages (Stages 3 and 4).

In a similarly designed study, Lan, Lucas, and Sun (2019) explored the association between NP complexity and L2 writing proficiency, using eleven nominal modifiers in Biber et al.'s (2011) index. The examined nominal modifiers demonstrated a significant association with proficiency differences, with four nominal modifiers (i.e., attributive adjectives, premodifying nouns, relative clauses, postmodifying *of* phrases) making the most salient contributions to this association. The post-hoc residual analysis further revealed that the more advanced group produced attributive adjectives (Stage 2) and relative clauses (Stage 3) more than expected, whereas the less advanced group produced premodifying nouns (Stage 3) and postmodifying *of* phrases (Stage 4) more than expected. The less proficient writers' heavy reliance on advanced noun modifiers seemed to contradict the developmental progression hypothesized by Biber et al. (2011). However, the qualitative analysis of those modifiers in the writing samples of the less advanced group revealed that a considerable number of premodifying nouns was affected by writing prompts, and most *of* phrases were associated with simple grammatical functions (e.g., *kind of*, *lot of*, *part of*).

Similar to Parkinson and Musgrave (2014) and Lan, Lucas, and Sun (2019), Kim (2020) investigated seven nominal modifiers based on Biber et al.'s (2011) hypothesized developmental stages. A total of 374 argumentative essays written by international undergraduate students from three different levels (i.e., low,

intermediate, upper-intermediate) were analyzed to determine whether adjacent levels of L2 writers can be distinguished by adopting the use of noun modifiers as discriminators for placement decisions. The results validated Biber et al.'s (2011) framework, showing that four nominal modifiers (i.e., premodifying adjective, prepositional phrase *of*, prepositional phrase other than *of* (abstract), multiple prepositional phrases) had a strong correlation with raters' placement decisions. The additional finding that the effects of the four nominal modifiers on raters' judgments were constant across two cumulative divisions of the three proficiency groups (i.e., low vs. intermediate and upper-intermediate, upper-intermediate vs. intermediate and low) bolstered the usefulness of the four nominal modifiers as robust discriminators for placement decisions. In terms of the relative contribution of the four noun modifiers to placement decisions, prepositional phrases *of* (Stage 3) were found to be most influential, whereas prepositional phrases other than *of* (abstract) (Stage 4) displayed the least impact. This result contradicted their corresponding developmental stages in Biber et al.'s (2011) model, but the researcher noted that this discrepancy could possibly be ascribed to (i) the range restriction due to the non-representative study population and (ii) timed argumentative essays, which may impede the use of advanced nominal modifiers.

Contrary to the general consensus of a total of five empirical studies examined thus far (i.e., Kim, 2020; Kyle & Crossley, 2018; Lan, Lucas, & Sun, 2019; Parkinson & Musgrave, 2014; Taguchi et al., 2013) on the significance of NP

complexity in advanced academic texts, some studies that explored the extent to which NP complexity predicts writing quality complicated this general picture. For instance, in Guo et al.'s (2013) study, two features pertaining to NP complexity (i.e., nominalizations, the number of modifiers per NP) were examined in relation to the quality of TOEFL writing tasks, and neither of them was reported as a reliable predictor of the task quality. Similarly, Crossley and McNamara (2014) examined clausal and phrasal complexity in essays of L2 students in a university ESL and EAP program to determine syntactic features predictive of essay quality. While acknowledging the nominal style of advanced writing, they also argued that the two NP-related features examined (i.e., number of modifiers per NP, subject relative clauses) were not highly predictive of human judgment of writing quality. Consistent with Guo et al. (2013) and Crossley and McNamara (2014), Yang et al. (2015), who examined the association between ESL graduate students' writing quality and syntactic complexity, reported that complex NPs per clause (CNP/C), which was the only NP-related feature analyzed in this study, had little predictive power for writing scores. As a similar line of domestic research, Park (2017) examined the distribution of NPs with different modifiers in argumentative essays of Korean college students and achieved mixed results for each type of NP. Specifically, attributive adjectives had no correlation with writing proficiency, and interestingly, premodifying nouns decreased at higher proficiency levels. In contrast, the two subtypes of NPs (i.e., relative clause, prepositional phrase other

than *of*) showed a slight increase, and the only area that exhibited a significant increase was prepositional phrase *of*.

Taken together, a wealth of empirical research yielded consistent findings in support of Biber et al.'s (2011) hypothesized progression from clausal elaboration to complex NPs, showing that phrasal modifiers embedded in NPs were positively associated with L2 writing proficiency. At the same time, though, conclusions reached in some studies were not compatible with Biber et al.'s (2011) argument, indicating that NP-related features had no significant correlation with writing quality. Such a disagreement points to the need to undertake more investigations to confirm whether the dense use of complex NPs reliably reflects L2 writing proficiency.

2.4 Variation in the Use of Grammatical Complexity

Features

As discussed in Section 2.3, studies that examined the association between grammatical complexity and L2 writing proficiency have reported contradictory findings. The inconsistency suggests the need to look at the influence of research design on the association. Not surprisingly, several learner-internal and learner-external factors come into play in connection with grammatical complexity of L2 writing. The present section is concerned with these factors, including L1 background (Section 2.4.1), genre (Section 2.4.2), and timing condition (Section

2.4.3).

2.4.1 The Effect of L1 Background

To date, most previous studies of L2 writing development were based on heterogenous L1 groups (e.g., Crossley & McNamara, 2014; Guo et al., 2013; Kim, 2020; Kyle & Crossley, 2018; Parkinson & Musgrave, 2014; Taguchi et al., 2013; Yang et al., 2015). In other words, these studies treated diverse L1 groups as a single homogenous group. Only one of the studies reviewed above (Lan, Lucas, & Sun, 2019) analyzed writing samples from the same L1 background (i.e., Chinese). As Lu (2011) noted, studies without considering L1 backgrounds “render(s) the reliability of the results obtained contingent on the untested assumption that the learner’s L1 does not significantly affect the relationship between syntactic complexity and language development” (p. 60). Lu (2011) further argued that inconsistent results for particular complexity measures, as outlined in the preceding section, are the corollary of research design differences.

Many empirical studies provided robust support for the notion that L1 background is associated with complexity features produced by L2 writers. These studies are rooted in the idea of cross-linguistic influence (CLI) (Sharwood Smith & Kellerman, 1986), which implies all potential influences of prior knowledge about other languages in the use of one language. For instance, Crossley and McNamara (2012) attempted to identify linguistic features related to specific L1

backgrounds by looking specifically into cohesion, lexical sophistication, conceptual knowledge, and syntactic complexity in English essays of L2 writers from four different L1 backgrounds. For syntactic complexity, only a single complexity measure (i.e., mean number of words before the main verbs) was employed, and the pairwise comparison displayed quite significant differences in complexity, showing that German writers used the largest number of words preceding the main verb, followed by Spanish, Finnish, and then Czech writers. These findings demonstrated “intragroup homogeneity” and “intergroup heterogeneity,” suggesting that L2 texts cannot be immune from writers’ L1 background (Crossley & McNamara, 2012, p. 123)

With the assistance of the L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010), Lu and Ai (2015) employed an entire set of fourteen syntactic complexity measures⁴ to evaluate syntactic complexity as a multifaceted construct. They analyzed English essays written by native speaker (NS) group and by non-native speaker (NNS) group of seven different L1 backgrounds. When treating all NNS groups as one combined group disregarding differing L1 backgrounds, significant differences were found in only three out of fourteen measures between the NS and

⁴ The measures in L2SCA are classified into five major types according to the associated complexity dimension: (a) length of production (mean length of clause [MLC], mean length of sentence [MLS], mean length of T-unit [MLT]), (b) sentence complexity (clauses per sentence [C/S]), (c) subordination (clauses per T-unit [C/T], complex T-units per T-unit [CT/T], dependent clauses per clause [DC/C], dependent clauses per T-unit [DC/T]), (d) coordination (coordinate phrases per clause [CP/C], coordinate phrases per T-unit [CP/T], T-units per sentence [T/S]), and (e) particular structures (complex nominals per clause [CN/C], complex nominals per T-unit [CN/T], verb phrases per T-unit [VP/T]) (Lu, 2011, p. 42).

NNS group (i.e., MLC, CN/C, CN/T). By contrast, when comparing the NS group with each of the seven NNS groups separately, all fourteen measures exhibited significant differences between the NS and more than one NNS group. Moreover, the patterns of difference were quite distinctive in each NNS group. These results suggest that treating NNSs with multiple L1 backgrounds as a single homogenous group can disguise the potential L1 influences on grammatical complexity of L2 writing. In other words, even L2 learners of similar proficiency levels can have different patterns in the use of complexity features due to the L1 factor. One interesting observation in this regard was made in the case of the Chinese group. Based on the Common European Framework of Reference (CEFR) proficiency levels, the Chinese group was rated as upper intermediate along with the Tswana and Japanese groups, while the other four NNS groups (i.e., German, French, Bulgarian, Russian) as advanced. More specifically, the Chinese group was the lowest ranked among all the NNS groups. When considering different sources of complexity according to L2 proficiency (Biber et al., 2011; Biber et al., 2020; Norris & Ortega, 2009; Ortega, 2012), the Chinese group was expected to achieve a lower level of phrasal complexity compared to clausal complexity. The results, however, turned out quite the opposite, and surprisingly, the Chinese writers were the sole NNS group that used considerably fewer sentential coordination than the NS group. This finding is linked directly with the characteristic of Chinese language that “the particular relationship between the two clauses is not signaled

explicitly and must be inferred by the hearer” (Li & Thompson, 1981, p. 641). Another noteworthy observation was that contrary to the expectation that advanced L2 writers produce relatively short and information-packed sentences with the dense use of phrasal modification, the German writers, the highest proficiency NNS group, produced significantly longer production units than the NS and the other NNS groups. This finding can be seen as reflecting the fact that “German sentences tend to be longer than English sentences. German sentences are usually more convoluted, and their structure tends to be hypotactic rather than paratactic” (Ziegler, 1991, p. 147). Taken together, it seems obvious that “proficiency-based prediction” does not always work, and L1 background should be considered as one important variable affecting the use of complexity features (Lu & Ai, 2015, p. 25).

In recognition of the lexis-grammar interrelation, Staples and Reppen (2016) investigated eight lexico-grammatical features produced by English L1 and L2 writers (i.e., Chinese and Arabic). They detected significant differences in four out of the eight features across three L1 groups (i.e., type/token ratio, premodifying nouns, complement clauses controlled by nouns, causative adverbial clauses), indicating the presence of considerable variation in syntactic patterns and their lexical realizations across different L1 backgrounds. Interestingly, L1 Chinese writers showed the highest degree of syntactic complexity in terms of premodifying nouns, followed by Arabic and then English writers, which was in

contrast with proficiency-based prediction. This finding is similar to Lu and Ai (2015) in that for L1 Chinese writers, despite their low proficiency, phrasal complexity appears to be relatively high. While the low lexical diversity due to the reuse of the same noun-noun sequences was found as one reasonable way to explain their greatest use of premodifying nouns, Staples and Reppen (2016) noted that further research is needed to reveal how L1 factors play a part in this phenomenon different than expectations.

Studies on Korean EFL learners' use of linguistic features have also shown possible L1 influence and Korean-specific preferences. Kang (2005), who examined cohesive devices and written discourse features produced by L1 Korean learners in their Korean and English written narratives compared to those produced by native English speakers (NESs) in their English narratives, observed that the linguistic features used by the Koreans did not show much difference in the two languages. For example, the Koreans produced considerably more conjunctive connectors (e.g., *so*, *therefore*) in both their L1 and L2 written narratives than the NESs' English narratives, which can be explained by L1 transfer. In other words, Korean learners' heavy reliance on conjunctions in English narratives may be attributed to "the relative clause-combining flexibility within the Korean language" (Kang, 2005, p. 275). In contrast, the Korean participants in this study seldom used syntactic features such as series (e.g., *No books, movies, or foods made her happy* rather than *No books made her happy, no*

movies made her happy, no foods made her happy) or multiple prepositional phrases (e.g., *There are some examples of the appropriate use of this word in sentences*) in both their L1 and L2 written discourse, suggesting that those are English-specific features which can pose a major challenge to Koreans. In sum, “they (L1 Korean learners) relied on their L1 writing skills to produce L2 written discourse, probably due to the challenges they have in actively using those English-specific features that are rarely used in their L1 written discourse” (Kang, 2005, p. 276). These findings provided empirical evidence for the influence of Korean language on Korean learners’ L2 writing and their difficulties in acquiring syntactic structures specific to English and absent in Korean.

Additional support for L1 transfer in Korean English learners’ use of linguistic features came from Park (2017), who investigated the distribution and internal structure of the NPs in writing samples of Korean college students. The researcher identified the rarity of participial pre- or post-modifiers in their writings (e.g., *a divided country, a country divided*) and associated the avoidance behavior with the “absence of this construction in L1 Korean” (p. 125). This interpretation was based on the contrasting results of Parkinson and Musgrave (2014) that participial premodifiers were not uncommon in the writings of international graduate students and their frequency was comparable to the frequency of relative clauses or premodifying nouns.

Insights from earlier studies of L1 influence on L2 written discourse point to

the importance of taking into consideration learners' L1 background when examining the relations between grammatical complexity and L2 proficiency. Therefore, it would not be reasonable to assume that the results on specific complexity measures obtained from studies using heterogeneous L1 groups are generalizable to specific L1 populations. In this regard, more empirical studies controlling the effect of L1 background are warranted. To successfully address potential L1-related factors and elucidate the developmental patterns of grammatical complexity distinctive of learners with a particular L1 background, researchers should analyze the data created by the target learner population only.

2.4.2 The Effect of Genre

Texts genres can be classified into two major categories: (1) narratives, working on “the description of events with a focus on people and their actions in a specific time frame,” and (2) non-narratives, centering around “making an argument or discussing ideas or beliefs in a logical fashion” (Berman & Slobin, 1994, as cited in Yoon & Polio, 2017, p. 280). From a functional perspective of language, the different communicative purposes of these two genres lead to differences in language use (Biber & Conrad, 2009; Halliday & Hasan, 1985).

The genre effect on the language produced by L2 writers has been extensively studied. For example, Way et al. (2000) examined three different genres of L2 French learners' writings (i.e., descriptive, narrative, expository), and found that

the expository task, which required composing a letter on American teenagers, for example, expressing their opinions on politics, produced the greatest syntactic complexity when measured via mean length of T-units. Besides, expository tasks proved to be the most challenging for L2 writers, showing the poorest performance in overall quality, fluency, and accuracy, respectively based on holistic scores, length of writing, and correct T-unit ratio.

Redressing the drawback of Way et al.'s (2000) study that only a single complexity measure was used, Lu (2011) examined the impact of two different genres (i.e., argumentative, narrative) on fourteen syntactic complexity measures by means of his L2SCA. Out of fourteen, thirteen measures exhibited a significant difference between argumentative and narrative essays, with the only exception of T/S, suggesting that argumentative essays produced greater syntactic complexity in general than narratives. Thus, the results of the two aforementioned studies consistently indicate that non-narrative genres elicit higher overall complexity than narrative genres.

More specifically, a body of empirical research has demonstrated that clausal complexity features are more prevalent in high-quality narratives than in non-narratives, whereas phrasal complexity features are more commonly used in high-quality non-narratives than in narratives. For instance, Beers and Nagy (2009), who investigated whether two different text genres (i.e., narrative, persuasive) influenced the relationship between syntactic complexity measures and writing

quality, reported that the writing quality of persuasive essays displayed a positive association with words per clause, but a negative association with clauses per T-unit. The quality of narratives showed the reverse, demonstrating no correlation with words per clause, but a positive correlation with clauses per T-unit. These findings suggest that syntactic complexity contributing to writing quality differs depending on text genre and that the kind of complexity associated with the quality of persuasive essays is clause-internal features typical of academic registers.

Similarly, Staples and Reppen (2016), who investigated the effects of two written genres (i.e., argumentative, rhetorical analysis) on fine-grained clausal and phrasal indices, reported that the writers, regardless of L1, used more attributive adjectives, nouns as premodifiers, and conditional adverbial clauses in argumentative writings than in rhetorical analysis. Given the greater use of all the phrasal features examined (i.e., attributive adjectives, nouns as premodifiers) in argumentative writings, this genre can be argued to have a typical characteristic of academic written discourse that heavily relies on phrasal modifiers.

Yoon and Polio (2017) confirmed Staples and Reppen's (2016) findings while partially disconfirming the findings of Lu (2011). To facilitate direct comparison, Yoon and Polio (2017) utilized the same automatic syntactic complexity analyzer used in Lu (2011) but excluded C/S and CT/T from the analysis as they proved to be not prominent development predictors (Ai & Lu, 2013; Lu, 2011). The results showed that argumentative essays exhibited greater complexity in terms of eight

out of twelve complexity measures than narrative essays. Notably, the four complexity measures that showed little genre effect (i.e., C/T, DC/C, DC/T, T/S) were all associated with clausal complexity such as subordination or coordination, whereas phrase-level complexity measures (i.e., CN/C, CN/T, VP/T) demonstrated a significant genre effect showing greater complexity in argumentative texts than in narratives.

As the review above suggests, different types of complexity features are required in different genres for high-quality texts. The different language use in narratives and non-narratives is a consequence of their different communicative purposes. In contrast to narratives in which “the vicissitudes of human intentions” are dealt with (Bruner, 1986, p. 16), non-narrative genres in which writers argue for a specific viewpoint require more use of phrasal complexity features to package vast quantities of information in a compact structure and to put their points across in a clear and concise way.

2.4.3 The Effect of Timing Condition

The impact of timing condition on language produced by L2 writers is well established. For instance, Ellis and Yuan (2004) explored the ways in which planning conditions on narrative writing influence learner output. In this study, planning conditions were designed in three types: no planning (NP), pretask planning (PTP), and on-line planning (OLP). Specifically, the NP condition

required a fast composition within the prescribed time limit, imposing restrictions on on-line planning. The PTP condition, on the other hand, allowed writers to plan for 10 minutes before they started writing, but as with the NP condition, they had limited chances for on-line planning with the prescribed time limit. Finally, in the OLP condition, unlimited time was allowed for writers to complete their tasks but no opportunity for pretask planning. The results of measuring syntactic complexity of L2 writers' production under these three types of planning conditions via clauses per T-unit (C/T) showed that the language produced by the two planning groups (i.e., the PTP and the OLP) were syntactically more complex than the NP group. Although little difference was detected between the two planning groups, the PTP group had a slightly larger effect size than the OLP group, suggesting that pretask planning contributes the most to the syntactic complexity of textual output.

Employing fourteen syntactic complexity measures included in L2SCA, Lu (2011) amply demonstrated that the degree of syntactic complexity was associated with timing conditions. To investigate the effects of timing condition, timed and untimed argumentative essays of L2 writers were analyzed based on these fourteen measures. Since the corpus examined in this study was built by nine different institutions and the topics of the essays differed for each institution, the impact of timing condition was investigated twice before and after controlling the institution. In the former case, untimed argumentative essays showed a greater degree of complexity than timed argumentative essays in terms of ten measures (i.e., MLC,

MLS, MLT, CT/T, DC/C, DC/T, CP/T, CN/C, CN/T, VP/T). After the institution was controlled, timing condition was found to exercise significant influence on seven syntactic complexity measures (i.e., MLC, MLS, MLT, CP/C, CP/T, CN/C, CN/T). Taken together, untimed argumentative essays, regardless of the topic effect, proved to be syntactically more complex than timed argumentative essays with respect to six measures (i.e., MLC, MLS, MLT, CP/T, CN/C, CN/T). Particularly noteworthy is that the two phrase-level complexity features related to extended NPs typical of formal academic writing (i.e., CN/C, CN/T) were all much more prevalent in untimed argumentative essays than in timed argumentative essays.

In sum, previous studies on the relationship of timing condition with grammatical complexity explicitly indicate that time pressure could negatively affect the complexity of language produced by L2 writers. In other words, L2 written production with a specific time limit might not reliably elicit complexity features typical of advanced academic writing such as phrasal features associated with complex NPs.

This chapter reviewed previous studies on grammatical complexity of L2 writing, discussing its importance as a discriminator of proficiency differences, its proper measurement methods, and lastly, a variety of factors that influence grammatical complexity. This literature review identified research gaps based on two principal findings. First, while many corpus-based studies following Biber et

al. (2011) concluded in agreement with their hypothesis of progressing from clausal to phrasal complexity, especially in NPs, some reported inconsistent results that complex NPs are not highly relevant to L2 proficiency. Second, despite the potential effects of L1 backgrounds on the use of complexity features in L2 writing, nearly all studies in this area used data from learners of mixed L1 backgrounds, which may be responsible for the aforementioned discrepancies in research results. These two findings illustrate the need to address L1 factors when examining the connection between L2 writing development and grammatical complexity. Thus, the current study analyzed the written production of English learners of L1 Korean background to determine whether the developmental patterns of learners with this specific L1 background conform to Biber et al.'s (2011) hypothesis for developmental progression. Specifically, three research questions are covered in this study: one examining an association between clausal and phrasal complexity in L1 Korean writers' essays and L2 writing proficiency, the second estimating the developmental trajectory of grammatical complexity by pinpointing specific complexity features as major contributors to the association, and the third qualitatively analyzing student writing to clarify the differences in the use of complexity features according to L2 writing proficiency.

CHAPTER 3. METHODOLOGY

This chapter introduces the research method of this study. First, the general profiles of the corpus used in the present study are outlined in Section 3.1. Next, the selection and classification of grammatical complexity measures included in the analysis are provided in Section 3.2. Then, the instrument for part-of-speech tagging and the process of extracting complexity features analyzed in this study are presented in Section 3.3. Lastly, the procedure of data analysis is explained in Section 3.4.

3.1 Learner Corpus

This study used a subset of the Yonsei English Learner Corpus (YELC 2011; Rhee & Jung, 2014). Subsection 3.1.1 provides the description of YELC 2011, and Subsection 3.1.2 contains the description of the corpus specifically built for this study.

3.1.1 Description of YELC 2011

YELC 2011, developed from 2011 to 2012 by Yonsei University in Korea, consists of 3,286 narrative and argumentative writings produced by 3,286 first-year students of Yonsei University in its English placement test (Yonsei English

Placement Test, YEPT) (Rhee & Jung, 2014). Every Yonsei University freshman is required to take YEPT. Under the exemption policy at this institution, however, those who are admitted to Underwood International University or who submit a high-performing certified English test report card are excluded from YEPT. The writing section of YEPT is subdivided into three parts. In Part 1, students are required to place the given words in the correct order. In Part 2, students are asked to write narrative essays on familiar topics related to their ordinary life within a maximum of 100 words; in Part 3, they are asked to write argumentative essays on academic subjects within 300 words. Among these three parts of YEPT, YELC 2011 was compiled by extracting writing samples from Part 2 and Part 3, which were renamed Part 1 and Part 2, respectively (Rhee & Jung, 2014). Table 3.1 summarizes the statistical description of YELC 2011.

TABLE 3.1
Statistical Information of YELC 2011

	Part 1	Part 2	YELC 2011
Texts	3,286	3,286	6,572
Tokens ⁵	315,317	770,511	1,085,828
Types ⁶	11,308	16,416	21,839
Standardized Type-Token Ratio ⁷	73.38	76.79	75.93
Total Number of Sentences	25,386	52,814	78,200
Number of Words per Sentence	12.36	14.57	13.85

Note. Adapted from Rhee and Jung (2014, p. 1025).

⁵ Token indicates “the total number of running words” in the texts (Scott, 2004, p. 214).

⁶ Type indicates the total number of “different words” in the texts (Scott, 2004, p. 154).

⁷ Standardized Type-Token Ratio (STTR), which complements the shortcomings of the text length-sensitive Type/Token Ratio (TTR), is “an average type-token ratio based on consecutive 1,000-word chunks of text” (Scott, 2004, p. 157).

The present study analyzed essays included in Part 2 (i.e., argumentative essays), and Table 3.2 provides details on the composition of the Part 2 dataset by proficiency. The essays were graded based on the holistic scale of CEFR, which describes the learner’s performance at six levels: A1, A2, B1, B2, C1, and C2.⁸ These six levels were subdivided into nine levels based on the nine-level scoring system of the Korean College Scholastic Ability Test (CSAT) to better represent the learner’s performance (Lee, 2011).

TABLE 3.2
Description of the Part 2 Dataset in YELC 2011

Proficiency	A1	A1+	A2	B1	B1+	B2	B2+	C1	C2
Text	41	185	684	1173	705	378	81	37	2
Token	1,544	22,100	133,675	276,414	191,967	108,404	24,358	11,411	629
Tokens per Text	37.7	119.5	195.4	235.6	272.3	286.8	300.7	308.4	314.5

Although the producer of YELC 2011 did not release the prompts presented to students, information on the writing topics used in Part 2 was reported in Choe and Song’s (2013) study. By adopting Hierarchical Agglomeration Clustering for topic classification, they categorized the 100 essays randomly selected from YELC 2011 into six topic groups: discipline, cellphone, smoking, animal, military, and Internet. With reference to these six keywords and the actual data of student

⁸ The CEFR comprises three major levels: A (basic), B (intermediate), and C (advanced). Each level is further broken into two levels resulting in a total of six levels that have a metaphorical name describing the language learners’ experience at a certain level: A1 (Breakthrough), A2 (Waystage), B1 (Threshold), B2 (Vantage), C1 (Effective Operational Proficiency), and C2 (Mastery) (McCarthy, 2016).

writing, six prompts were estimated as (a) allowing physical punishment in schools, (b) using cellular phones while driving, (c) prohibiting smoking in public areas, (d) experimenting on animals for medical purposes, (e) Korea's policy of compulsory military service, and (f) the real-name policy for the Internet. The current study analyzed randomly extracted data from these six topics, so the possible impact of prompts on grammatical complexity could not be controlled.

3.1.2 Description of a Subset of YELC 2011 used in the Study

Given the significant impact of genre on the values of grammatical complexity measures (Beers & Nagy, 2009; Lu, 2011; Staples & Reppen, 2016; Way et al., 2000; Yoon & Polio, 2017), this variable needed to be held constant so as not to affect the relationship of L2 writing proficiency with grammatical complexity. Thus, as previously noted, only argumentative essays (Part 2) were selected for this study to control for genre effects. The decision was made based on the finding of preceding research that argumentative essays elicit linguistic features typical of academic prose more than narrative essays. Similar findings were yielded in Park's (2017) work, which investigated the impact of genres using YELC 2011 and showed that argumentative essays produced more NPs with modifiers than narratives. Her study consistently suggests that argumentative genre can be considered more equivalent to formal academic writing in terms of communicative

purposes and associated language use compared to narratives.

Based on the similarity of argumentative essays to academic genre, this study retrieved a total of 234 argumentative essays at seven out of nine different levels of writing proficiency from YELC 2011 Part 2 (see Table 3.3). The two borderline levels A2 and B2 were excluded to guarantee disparities in proficiency between groups. Since all student writing samples in this corpus are divided into subgroups based on CEFR writing proficiency levels, stratified random sampling could be employed to ensure representation of each level (Acharya et al., 2013). Considering that the total number of C1 and C2 level essays was 39, the same number of essays were extracted from the other five levels. By categorizing these seven levels into three levels, the corpus for this study was comprised of three subcorpora representing the low (A1 and A1+), intermediate (B1 and B1+), and high levels (B2+, C1, and C2) of L2 writing proficiency with an attempt to observe the developmental trajectory of grammatical complexity of Korean college students.

TABLE 3.3
Description of the Selected Data

Subcorpora	Low		Intermediate			High	
	A1	A1+	B1	B1+	B2+	C1	C2
Proficiency Text	39	39	39	39	39	37	2
Total Number of Texts	78		78			78	
Total Tokens	6334		19554			23861	
Tokens per Text	81.2		250.7			305.9	

3.2 Grammatical Complexity Measures

Following the developmental progression from clausal subordination to complex NPs hypothesized by Biber et al. (2011), this study analyzed grammatical complexity at two levels: clause and phrase levels. Table 3.4 lists individual complexity features under analysis. Under clausal structures, there were three major subtypes: finite adverbial clauses, finite complement clauses, and finite noun modifier clauses. There were more distinctions within each subtype. The finite adverbial clauses further included three specific subcategories: causative clauses (*because*), conditional clauses (*if*), and concessive clauses (*although*). As in Biber et al.'s (2011) study, these three specific adverbial subordinators were chosen, because they can be differentiated from other subordinators (e.g., *since*, *as*, *while*) that are commonly used in academic prose rather than in conversational discourse due to ambiguity for interpretation caused by their multiple meanings (Biber, 2006). There were also four specific types of finite complement clauses: *that*-clauses headed by verbs, *wh*-clauses headed by verbs, *that*-clauses headed by adjectives, and *that*-clauses headed by nouns. The finite noun modifier clauses included *that* relative clauses and WH relative clauses (*who*, *which*, *whom*, *whose*).

Phrasal structures, on the other hand, included three major subtypes of nominal modifiers: premodifying adjectives (attributive and participial), premodifying nouns, and postmodifying prepositional phrases (*of*, *in*, *on*, *with*, *for*).

TABLE 3.4
Clause- and Phrase-level Complexity Measures used in the Analysis

Type	Stage	Linguistic features	Examples	
Clause-level complexity measures	2	Finite adverbial clauses		
		Causative	She won't narc on me, <u>because she prides herself on being a gangster.</u>	
		Conditional	<u>If I stay here,</u> I'll have to leave early in the morning.	
	2	Concessive	If I don't put my name, she doesn't know who wrote it, <u>although she might guess.</u>	
	1-2	Finite complement clauses		
		Verb + <i>that</i> clause	I would hope <u>that we can have more control over them.</u>	
		Verb + WH clause	I don't know <u>how they do it.</u>	
		3	Adjective + <i>that</i> clause	I was sure <u>that I could smooth over our little misunderstanding.</u>
		5	Noun + <i>that</i> clause	The fact <u>that no tracer particles were found in or below the tight junction</u>
	3	Finite noun modifier clauses		
<i>That</i> relative clause		Experimental error <u>that could be expected to result from using cloze tests</u>		
?	WH relative clause	A ring <u>which limits a central electron transparent space</u>		
Phrase-level complexity measures		Noun modifiers		
	2	Premodifying adjective:		
		Attributive	<u>Emotional</u> injury	
	?	Participial ^a	<u>Contaminated</u> world (<i>-ed</i> participial) <u>Devastating</u> tsunami (<i>-ing</i> participial)	
	3	Premodifying noun	The <u>trial transfer</u> sessions	
	3-5	Prepositional phrase as postmodifier:		
	<i>Of</i> as postmodifier	Editor <u>of the food section</u>		
	<i>In</i> as postmodifier	House <u>in the suburbs</u>		
	<i>On</i> as postmodifier	Significant effects <u>on our perceptions</u>		
	<i>With</i> as postmodifier	Sociology <u>with a system, theoretical orientation</u>		
	<i>For</i> as postmodifier	Scores <u>for male and female target students in the class</u>		

Note. Adapted from Biber et al. (2011). The examples are from Biber (2006), Biber et al. (2011), and Parkinson and Musgrave (2014).

^a Not included in Biber et al. (2011).

Given the developmental progression hypothesized by Biber et al. (2011), phrase-level complexity features generally represent a higher level of complexity characteristic of academic writing, whereas clause-level complexity features are normally expected to have useful predictive power at developmentally less advanced levels. Thus, the combined use of clause-level and phrase-level measures was pursued in the current study to unveil a complete picture of developmental patterns across proficiency levels (Norris & Ortega, 2009). The complexity features at the intermediate stages of Biber et al.'s (2011) index such as phrases functioning syntactically as clause constituents and nonfinite clause types were eliminated from this study since they turned out to be not statistically prominent complexity features that distinguish academic prose from conversation.

The second column of Table 3.4 shows the developmental stage for each complexity feature suggested by Biber et al. (2011). As previously shown in Table 2.2, verb complement clause is placed in Stages 1-2 of Biber et al.'s (2011) index depending on the kinds of verbs controlling *that*-clauses. Besides, it is important to note that noun complement clause, despite its status as a clausal structure, is assumed to be acquired at the final stage of development (Stage 5). Unlike other clausal features, this structure is commonly used in academic writing to convey stance meanings, and it functions syntactically as nominal modifier formulating extended NPs typical of formal written registers (Biber, 2006; Biber et al., 2011; Staples & Reppen, 2016). Thus, the findings on this construction should be

interpreted differently than other clausal complexity features. The developmental stage to which WH relative clause belongs is not explicitly listed in Biber et al.'s (2011) model. However, considering that WH relative clause was found to be much more common in academic prose than in conversation unlike *that* relative clause (Biber et al., 2011), one might tentatively infer that WH relative clause would be acquired developmentally later than *that* relative clause. The stage where participial adjective as nominal premodifier is acquired is also not stated in Table 3.4, because this feature was not examined in Biber et al. (2011). The present study additionally analyzed participial premodifiers since previous studies provided empirical evidence that they showed a statistically significant difference among L2 writers with different proficiency levels (Kim, 2020; Parkinson & Musgrave, 2014). The two subtypes of participial adjective (i.e., *-ed* participial, *-ing* participial) were analyzed as illustrated in Table 3.4. Postmodifying prepositional phrases are located in Stages 3-5 of Biber et al.'s (2011) framework. As described previously in Table 2.2, prepositional phrases *of* and other than *of* with concrete/locative meanings are postulated for Stage 3, prepositional phrases other than *of* with abstract meanings for Stage 4, and prepositional phrases with nonfinite complement clauses and multiple prepositional phrases for Stage 5.

In terms of phrase-level complexity measures, an NP that has a pronoun as its head was excluded in this study as most pronouns are not accompanied by complements or modifiers except in a few cases (Biber et al., 1999). Considering

these exceptional cases, only the three types of pronoun-headed phrases, namely, NPs headed by pronouns such as *one*, *that*, and *those* were included (e.g., *non-smoking one*, *that of Korea*, *those in medical experiments*). In addition, nominal gerund phrases occurring in the place of subjects, subjective predicative, or direct objects were included since they can be analyzed based on the relationship of the head noun-modifier(s) (e.g., *working for industrial development*, *banning smoking in public buildings*). On the other hand, proper nouns used as nominal modifiers were excluded from the analysis as they were seen as simply listing words rather than reflecting a student's writing proficiency (e.g., *USA armies*, *Hae-Byung-Dae applicants*, *Seoul City*).

3.3 Corpus Tagging and Automatic Extraction

Student essays from YELC 2011 were annotated by automatic part-of-speech (POS) tagging software called CLAWS (the Constituent Likelihood Automatic Word-tagging System) (see Appendix 1 for a sample tagged text). The latest version, CLAWS4, known as the British National Corpus (BNC) tagger, works with a high accuracy rate of 96-97% with the variation of the actual rate depending on the text type, as a hybrid grammatical tagger that combines both probabilistic and rule-based procedures (Garside & Smith, 1997). The probabilistic approach is used to select a tag for a word with the highest probability using the context of the word, and the rule-based approach is incorporated to address “the problem of how

to deal with idiosyncratic word sequences or multiword like *as well as*” (Garside & Smith, 1997, p. 105). Despite CLAWS being trained on native speaker corpora, studies that tested the performance of CLAWS on learner corpora reported an accuracy rate equal to or slightly lower than that achieved on the BNC, for example achieving 96% accuracy on the Tswana Learner English Corpus (TLEC) (Van & Schäfer, 2002) and 93.6% accuracy on the Malaysian Corpus of Learner English (MACLE) (Aziz & Don, 2019). Thus, CLAWS was chosen to POS tag the dataset analyzed in this study.

Then, the extraction of the 17 linguistic features in Table 3.4 from the tagged texts was performed through the two-step process: (1) automatic extraction and (2) visual inspection. The first step was to automatically extract linguistic features based on the tagged corpus using regular expressions (REGEX), which mean “simple or very complicated sequences of characters in files” (Gries & Newman, 2013, p. 270). The CLAWS tagset was used to generate REGEX for each complexity feature analyzed in this study. The tagset used in this study was C7, which contains 140 tags.⁹ By means of REGEX containing POS information, concordance lines with the target features could be obtained in text editor Notepad++ (see Appendix 2 for concordance lines of attributive adjectives). It should be noted that omitted grammatical features cannot be identified by the CLAWS tagger. Thus, the analysis included only those instances where a

⁹ The C7 tagset is available at <https://ucrel.lancs.ac.uk/claws7tags.html>

complementizer in *that*-clauses or a relativizer in relative clauses is present.

Next, in the process of visual inspection, the researcher read through each concordance line to enhance the accuracy rate since no automatic extraction ensures accurate identification of target features. To illustrate, the automatic analysis failed to distinguish between *that* relative clauses and *that* complement clauses and between prepositional phrases as an adverbial versus a postnominal modifier. Except for *of* phrases, which mostly function as nominal postmodifiers when occurring immediately after nouns, other prepositions (*in, on, with, for*) should be manually checked to determine their syntactic function. In cases where a prepositional phrase was syntactically ambiguous and could be interpreted either as an adverbial phrase or as a postnominal modifier, it was excluded from the analysis following Biber et al. (2011). With respect to *of* phrases, those following “partitive nouns” (Quirk et al., 1985, p. 249) or “collective nouns” (e.g., *group of*), “unit nouns” (e.g., *bit of*), “quantifying nouns” (e.g., *couple of*), and “species nouns” (e.g., *kind of*) (Biber et al., 1999, p. 247-257) were omitted from the study as well. In these constructions, it is not always clear to determine whether the head of an NP is the noun following the preposition *of* or preceding it, so *of* phrases with these package nouns¹⁰ are “not true instances of postmodification” (Biber et al., 1999, p. 636). In addition, package nouns are likely to show distinct

¹⁰ Package nouns are the collective name for unit nouns, collective nouns, species nouns, quantifying nouns, which perform the common functions of “packaging together a range of entities” (Biber, Conrad, & Leech, 2002, p. 60).

collocational patterns and primarily concerned with interpersonal spoken register rather than academic register (Biber et al., 1999). Thus, Lan, Lucas, and Sun (2019) reported that less proficient writers are more likely to resort to *of* phrases with nouns of this kind due to their simple grammatical functions. However, this is not the case for two species nouns: *type(s) of* and *species of*. In contrast to other species nouns such as *sort(s) of* and *kind of*, they are predominantly associated with academic prose since “classification is an important aspect of academic procedure and discourse” (Biber et al., 1999, p. 256). Accordingly, these two species nouns (i.e., *types(s) of*, *species of*) were included in the analysis, with all other package nouns ruled out.

With respect to finite complement clauses headed by a verb or an adjective, only post-predicate complement clauses (e.g., *Some people say that it is violation of animal rights*) were considered, excluding extraposed ones from the study (e.g., *It seemed that there were no problems*)¹¹ based on the distributional patterns of these two *that*-clause types across registers. Post-predicate *that*-clauses are primarily associated with conversational discourse rather than academic prose due to their characteristics particularly suited for interpersonal communication such as “a personal, human noun phrase as subject of the main clauses, and an active voice

¹¹ Three main grammatical positions of complement clauses are as follows: pre-predicate, post-predicate, and extraposed positions (Biber, Conrad, & Leech, 2002). The pre-predicate position refers to when complement clauses occur before the verb or adjective, whereas the post-predicate position refers to when complement clauses occur after the verb or adjective. Extraposed position is an alternative to the pre-predicate position, in which *that*-clause in the subject position is moved to the post-predicate position, and dummy pronoun *it* fills the subject slot.

verb as the predicate” (Biber et al., 1999, p. 674). Extraposed *that*-clauses exhibit the opposite register distribution with contrasting characteristics such as “an impersonal, non-human noun phrase as subject of the main clause” and “a main clause predicate representing a static relation or attribute” (Biber et al., 1999, p. 675). Accordingly, these structures were hypothesized as being acquired later (Stage 4) than post-predicate complement clauses after verbs (Stages 1 and 2) or after adjectives (Stage 3) by Biber et al. (2011). Despite being based on Biber et al. (2011), Taguchi et al. (2013) classified these two types of complement clauses into the same level of complexity measures and failed to consider differences in their functions and associated registers. This could be a reason for the results of their study that verb complement clauses were found to be much more used by advanced than less advanced L2 writers and that adjective complement clauses showed no significant difference. In this study, given that complement clauses headed by a verb or an adjective were analyzed to tap grammatical complexity at lower levels of development in conjunction with other clause-level complexity measures, only post-predicate complement clauses were subject to statistical analysis. Extraposed complement clauses, on the other hand, were removed by hand as a high level of complexity indicator unlike other finite complement clauses.

Another case that needed visual inspection was the use of multiple adjectives or nouns as nominal premodifiers. Regardless of whether they are coordinated premodifiers (e.g., *dangerous and unstable conditions*) or not (e.g., *alternative*

military service), the use of consecutive premodifiers was recognized as a single modifier by automatic extraction, so it needed to be accurately calculated through visual inspection. Additionally, the researcher manually handled the following two cases for all the target features: (a) the erroneous extraction due to misleading tags caused by learner errors was removed from the output; and (b) lexical bundles (e.g., *in other words, on the other hand, point of view, a series of*) and phrasal verbs (e.g., *take care of, take advantage of, take control of*), which are used as prefabricated expressions, were also excluded by hand.

To check the accuracy rate of automatic extraction and visual inspection, the researcher manually coded the occurrences of all the target complexity features in randomly selected 30 writing samples (approximately 10% of the total texts) and then compared the outcome of the manual work to that of automatic extraction accompanied by visual inspection. Of the total seventeen complexity features, the accuracy rates of thirteen were 100%.¹² Of the remaining four, three were about 95%, with the top rate for attributive adjectives (99.5%) followed by WH relative clauses (98.0%) and *that*-clauses headed by verbs (94.2%). *Wh*-clauses headed by verbs showed the lowest accuracy rate (85.0%). For these four complexity features, the failure of automatic extraction with REGEX occurred when the components of the target features were separated by cardinal numbers (e.g., *following two*

¹² The cases where the target features failed to be auto-extracted due to incorrect tagging were not included in the calculation.

reasons), prepositional phrases (e.g., *show to the world that*), or coordinating conjunctions (e.g., *students who... and who*). The lowest accuracy rate of *wh*-clauses headed by verbs was largely due to their relatively low frequency of occurrence. These features were extremely rare in student writing in absolute terms, so despite the number of extraction failures being either similar to or even fewer than other features, the accuracy rate of these features was calculated inevitably as the lowest. The overall accuracy rate for all the complexity features (i.e., 99.2%) was considered high enough for subsequent data analysis.

3.4 Data Analysis

To address the research questions stated in Section 1.3, both quantitative and qualitative data analyses were conducted in the current study. In other words, quantitative findings based on frequency counts of complexity features for each proficiency level were checked again by qualitatively reviewing the actual use of those features in student writing.

For quantitative analysis, two statistical tests were applied in *Statistical Package for Social Sciences 26* (SPSS): a Pearson Chi-square test and a residual analysis.¹³ The Chi-square test provides several advantages in terms of “its robustness with respect to distribution of data, its ease of computation, the detailed

¹³ The combination of these two statistical tests to investigate grammatical complexity of L2 writing was first used by Lan, Lucas, and Sun (2019).

information that can be derived from the test,” and “its flexibility in handling data from both two group and multiple group studies” (McHugh, 2013, p. 143). More important, it can be used to check “if frequencies on one variable (the dependent variable) change with levels of another independent variable” (Hatch & Lazaraton, 1991, p. 399). This study attempted to test if the frequencies of complexity features (the dependent variable) change with proficiency levels (the independent variable), or if the use of the nine complexity features is associated with the three proficiency levels. To that end, the Chi-square test was conducted based on a 9 x 3 contingency table (i.e., a total of 27 cells¹⁴) with 16 degrees of freedom. As presented in Section 3.2, the 17 complexity measures were originally set up for the test, but they were combined into the nine measures based on similarities in grammatical structures. The category collapsing was done to ensure enough cell frequencies for inferential statistics (i.e., the Chi-square statistic) considering its assumption that cells with an expected frequency¹⁵ of less than 5 should not exceed 20% of the total (McHugh, 2013). For example, as the observed frequency of adjective complement clauses in this study was less than 3 for all proficiency levels, statistical analysis was applied after combining it with the frequency of verb complement clauses based on their structural similarity. For the same reason, three

¹⁴ In this study, a cell denotes a specific complexity feature produced by a specific proficiency level, for example, premodifying nouns used by low-proficiency students.

¹⁵ The expected values represent the estimated rate of occurrence of the complexity features if there were no connection between L2 writing proficiency and the use of complexity features. To obtain the expected value for each cell, “its row marginal is multiplied by its column marginal, and that product is divided by the sample size” (McHugh, 2013).

types of finite adverbial clauses (i.e., causative, conditional, concessive), two types of premodifying adjectives (i.e., attributive, participial), and prepositional phrases other than *of* (i.e., *in*, *on*, *with*, *for*) were each collapsed into a single category for a Chi-square test.

Given that “the Chi-square test is a significance statistic and should be followed by a strength statistic” (McHugh, 2013, p. 143), Cramer’s V was performed in tandem to reveal “the strength of the association” (ibid., p. 148). It ranges from 0 to 1, with values nearer to 1 showing larger effects. Table 3.5 summarizes how to interpret Cramer’s V. It is basically a type of correlation (McHugh, 2013), so its high value in this study will indicate that grammatical complexity of L2 writing depends greatly on writers’ proficiency.

TABLE 3.5
Interpretation of Cramer’s V

Cramer’s V	Interpretation
> 0.25	Very strong
> 0.15	Strong
> 0.10	Moderate
> 0.05	Weak
> 0	No or very weak

Note. From Akoglu (2018, p. 92).

Additionally, a residual analysis was performed as follow-up probing to address “the omnibus nature of the Chi-square tests” (Sharpe, 2015, p. 1). The Chi-square value only provides information on whether there is a significant association, but the source of significant results is not evident. One approach to

deal with the holistic test problem is calculating residuals. A residual refers to “the difference between the observed and expected values for a cell,” and “the larger the residual, the greater the contribution of the cell to the magnitude of the resulting chi-square obtained value” (Sharpe, 2015, p. 2). Thus, a residual analysis was undertaken in this study to identify which complexity features, or which cells among a total of 27 cells in the contingency table, influenced the overall significant association the most. To be more precise, adjusted standardized residuals, rather than raw residuals, were used, given the effect of cell size. In other words, adjusted standardized residuals were selected to avoid the redundancy problem that “cells with the largest expected values also produce the largest raw residuals” (Sharpe, 2015, p. 3). Agresti (2018) suggested that “a standardized residual having an absolute value that exceeds about 2 when there are few cells or about 3 when there are many cells indicates lack of fit of H_0 in that cell” (p. 39). Following this rule of thumb in statistics, Lan et al. (2022), who explored grammatical complexity in L2 written production, adopted ± 2 criteria when identifying cells with large residuals. Accordingly, this study considered cells with adjusted residuals greater than $|2.0|$ as having large residuals worth proceeding with qualitative analysis in which concordance lines for target complexity features were scrutinized in terms of their lexical realizations and accuracy.

The qualitative review was based on the idea that greater use of particular grammatical features may not always indicate greater development. For a more

accurate estimation of writing development, lexical diversity, academically oriented lexical choices, and the accuracy with which these features are used should be evaluated together, beyond the frequency-based perspective. Thus, the actual use of complexity features in student writing was qualitatively analyzed to gain better insights for academic writing development and to better interpret quantitative findings on the association between grammatical complexity and L2 writing proficiency.

CHAPTER 4. RESULTS AND DISCUSSION

This chapter reports and discusses research findings. First, descriptive statistics for the 17 complexity features (i.e., the nine clause-level and eight phrase-level complexity features) are presented in Section 4.1. Then, Section 4.2 examines how the use of complexity features is associated with L2 writing proficiency based on the results of the Chi-square test and Cramer's V to address the first research question. Next, Section 4.3 suggests the developmental patterns of grammatical complexity based on the results of residual analysis that reveal specifically which features made significant contributions to the association as an answer to the second research question. Lastly, Sections 4.4 and 4.5 provide more accurate interpretations of the quantitative findings by qualitatively analyzing the use of complexity features in essay samples to answer the third research question.

4.1 Descriptive Statistics

The frequency for each grammatical complexity feature by proficiency is provided in Table 4.1. The normed frequency in parentheses¹⁶ can be used for a direct comparison of the three proficiency groups from the frequency perspective by eliminating the effect of text length (Biber, 2006).

¹⁶ The formula used to calculate the normalized frequency for each complexity feature was (raw frequency/total number of words) *1,000. The total word counts for low-, mid-, and high-rated essays were 6,334, 19,554, and 23,861 words.

TABLE 4.1
Frequencies of Grammatical Complexity Features

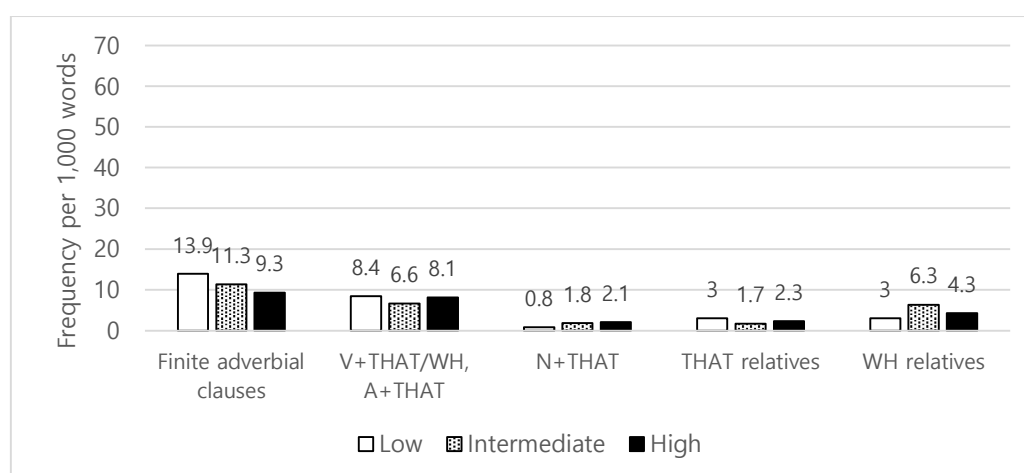
Type	Grammatical structures	Low	Intermediate	High
Clause-level complexity	Causative (<i>because</i>)	45 (7.1)	82 (4.2)	88 (3.7)
	Conditional (<i>if</i>)	38 (6.0)	132 (6.8)	113 (4.7)
	Concessive (<i>although</i>)	5 (0.8)	7 (0.4)	20 (0.8)
	Verb + <i>that</i> clause	43 (6.8)	103 (5.3)	148 (6.2)
	Verb + WH clause*	10 (1.6)	25 (1.3)	42 (1.8)
	Adjective + <i>that</i> clause	0 (0)	2 (0.1)	3 (0.1)
	Noun + <i>that</i> clause*	5 (0.8)	35 (1.8)	50 (2.0)
	<i>That</i> relative clause	19 (3.0)	33 (1.7)	55 (2.3)
	WH relative clause	19 (3.0)	123 (6.3)	103 (4.3)
Phrase-level complexity	Attributive as premodifiers*	383 (60.5)	1122 (57.4)	1450 (60.8)
	Participial premodifiers	21 (3.3)	69 (3.5)	81 (3.4)
	Nouns as premodifiers*	43 (6.8)	126 (6.4)	196 (8.2)
	<i>Of</i> phrases as postmodifiers*	42 (6.6)	170 (8.7)	336 (14.1)
	<i>In</i> phrases as postmodifiers*	4 (0.6)	32 (1.6)	42 (1.8)
	<i>On</i> phrases as postmodifiers*	3 (0.5)	16 (0.8)	26 (1.1)
	<i>With</i> phrases as postmodifiers	21 (3.3)	74 (3.8)	88 (3.7)
<i>For</i> phrases as postmodifiers*	4 (0.6)	18 (0.9)	28 (1.2)	

Note. The numbers in parentheses indicate the normed frequency per 1,000 words. The asterisk (*) indicates that the high-rated essays exhibit the highest normed frequency of a given complexity feature.

Figures 4.1 and 4.2 graphically show the comparison of the normed frequencies after category collapsing. Figure 4.1 plots frequency per 1,000 words for clause-level complexity features, showing that three out of five clausal features (i.e., finite adverbial clause (*because, if, although*), finite complement clause headed by verbs or adjectives, *that* relative clause) were most common in the low-rated essays, whereas one clausal feature (i.e., finite complement clause headed by nouns) was most common in the high-rated essays. When it comes to finite adverbial clause and finite clause functioning as noun complement, a gradual decrease or increase in frequency occurred according to proficiency levels. WH

relative clause was clearly distinguished from all the other clausal features by its highest frequency in the mid-rated essays. The overall results suggest that the low- and mid-proficiency groups generally used clausal features at a greater frequency than the high-proficiency group but in different ways: the low-rated essays were primarily associated with finite adverbial clauses, whereas the mid-rated essays were mainly related to WH relative clauses.

FIGURE 4.1 Frequencies of Clause-level Complexity Features

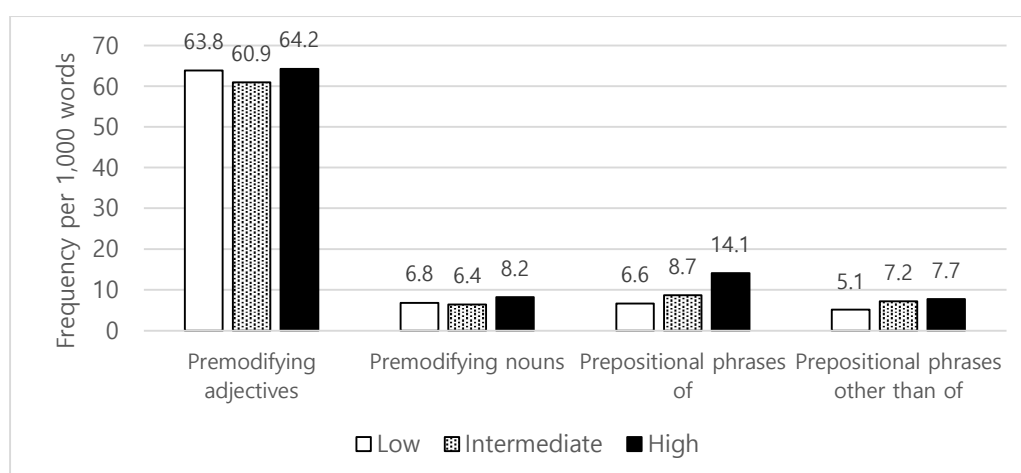


In stark contrast, all the phrase-level complexity features were most prevalent in the high-rated essays as shown in Figure 4.2. The most striking difference in frequency was found in *of* phrase as a nominal postmodifier. Figure 4.2 also suggests that as opposed to nominal premodifiers (i.e., premodifying adjective¹⁷ and noun), nominal postmodifiers (i.e., prepositional phrase *of* and other than *of*)

¹⁷ In this study, premodifying adjectives include both attributive and participial adjectives.

exhibited a gradual increase in frequency towards the higher proficiency levels, tentatively suggesting that postmodification provides better indicators of advanced proficiency levels than premodification. The learners' L1 background seems to give one possible reason for this finding because Korean is a head-final language that does not allow noun phrases to be extended with the addition of postmodifiers (Baik, 1994).

FIGURE 4.2 Frequencies of Phrase-level Complexity Features



In sum, it seems as though the low- or mid-rated essays could be considered more grammatically complex than the high-rated ones when taking into account clausal complexity. However, markedly different results could be achieved when evaluating complexity via phrasal elaboration. Examples (1) through (3) demonstrate different kinds of grammatical complexity preferred by each proficiency level; subordinators and *wh*-words are in bold with the dependent

clauses they introduce underlined; premodifying adjectives are in italics; postmodifying prepositional phrases are in bold italics.

- (1) I agree with them at first, **because** I was in school with a lot of *trouble-making* guys, I cannot concentrate fully on my own textbook. <Low, file 2609>
- (2) **If** accidents happen because of using *cellular* phones while driving, victims **who** get accident are not limited to drivers **who** use *cellular* phones while driving. <Mid, file 2849>
- (3) First of all, the *main* assertion **of people who** speak of completing *military* service as a waste *of time* cannot be agreed. <High, file 32>

Example (1) from a low-rated essay has one clausal complexity feature (*because* clause) and one phrasal complexity feature (participial adjective *trouble-making*) in a 25-word sentence. Examples (2) and (3) from mid- and high-rated essays, despite being of a similar length (25 and 23 words, respectively), have more complexity features. Example (2) includes three clausal complexity features (one *if* clause, two WH relative clauses) and two phrasal features (attributive adjective *cellular*).¹⁸ Example (3) is obviously more complex than Examples (1) and (2) with respect to phrasal complexity, including two attributive adjectives (*main*, *military*) and two postmodifying prepositional phrases (*of people*, *of time*) as well as one clausal feature (WH relative clause). One particularly interesting observation is that all the complexity features used in Example (3) function

¹⁸ The count of complexity features is from the frequency perspective without consideration for lexical diversity.

syntactically as NP constituents, which results in a heavy, complex NP containing plenty of concise information in the subject position of the sentence.

The differences in frequency distributions of the clause- and phrase-level complexity features across proficiency levels suggest that L2 writers' proficiency is potentially relevant to grammatical complexity. In general, student writers showed a tendency to move from using clause-level complexity features at lower levels to using phrase-level complexity features at upper levels. These patterns of use are comparable to Biber et al.'s (2011) argument that "the stages generally progress from finite dependent clauses functioning as constituents in other clauses" and "finally to the last stage requiring dense use of phrasal (nonclausal) dependent structures that function as constituents in noun phrases" (pp. 29-30). Accordingly, the high-rated essays, in which the role of dependent clauses subsided in favor of phrasal noun modifiers, can be argued to show advanced grammatical complexity characteristic of mature academic writing. This interpretation, however, needs further statistical evidence, given that the normalized frequencies of phrasal features in highly rated essays are not exceptionally high, and thus, their role as an indicator of L2 writing quality seems still tentative. Therefore, the association between the use of complexity features and L2 writing proficiency is statistically analyzed in the following sections.

4.2 The Association between L2 Writing Proficiency and Grammatical Complexity

The association between two nominal variables (i.e., L2 writing proficiency and clause- and phrase-level complexity features) was investigated by applying Pearson Chi-square (χ^2) test. The SPSS output is presented in Table 4.2. The calculated χ^2 value is 82.478 based on 16 df and p value is .000. This indicates an association between the nine complexity features and L2 writing proficiency at the significance level of .001.

TABLE 4.2
Results of the Pearson Chi-square Test and Cramer's V

	value	DF	p value	Cramer's V
Pearson Chi-square	82.478	16	.000	.085

Note. DF = degrees of freedom.

Cramer's V for strength testing is .085, which means that the use of the nine complexity features accounts for 8.5% variance of L2 writing proficiency. According to the conventions (see Table 3.5), Cramer's V of less than 0.10 indicates a weak association. The small effect size based on Cramer's V value (.085) is a predictable outcome in a research context where a limited set of fine-grained indices are employed "in a complex register (i.e., academic writing) to investigate their association with a complex construct (i.e., writing proficiency)" (Lan, Lucas, & Sun, 2019, p. 8). A larger effect size could be generated using large-

grained measures¹⁹ based on the cumulative effects of several fine-grained measures involving clausal, phrasal, and lexical features while admitting the multicollinearity problem among variables (Lan et al., 2022). Besides, it has been well documented in preceding corpus-based studies that writing quality can be better determined by looking at how a wide variety of linguistic features co-occur rather than examining any isolated linguistic features because they are motivated by related communicative functions (Biber et al., 2016; Friginal et al., 2014). The notion of linguistic co-occurrence provides a possible interpretation of the small effect size obtained in this study by implying that instructors or raters are “much more tuned in to constellations of linguistic features” with a priority on overall communicative effectiveness (Biber et al., 2016, p. 657). Given the holistic nature of writing scores assigned by raters, the result that only nine grammatical categories analyzed in this study can account for 8.5% variance of differing proficiencies can be interpreted as a significant effect.

4.3 The Developmental Patterns of Grammatical Complexity

To explore further the statistically significant holistic value obtained in the

¹⁹ Large-grained measures index complexity at the sentence or clause level (e.g., length of any production unit, amount of subordination), whereas fine-grained measures primarily rely on various types of subordinate clauses and nominal modifiers (Jiang et al., 2019; Kyle & Crossley, 2018).

Chi-square analysis, a residual analysis was applied as a follow-up testing. Results of the residual analysis pinpoint which specific cells, or grammatical complexity features, made significant contributions to the magnitude of the obtained Chi-square value. Table 4.3 presents the adjusted standardized residuals of the nine grammatical complexity features.

TABLE 4.3
Results of the Adjusted Standardized Residuals

	Low	Intermediate	High
Clause-level complexity features			
Finite adverbial clause*	3.2**	2.0	-4.0
Verb/adjective complement clause	1.1	-1.3	0.6
Noun complement clause*	-2.0	0.2	1.1
<i>That</i> relative clause	1.7	-1.5	0.3
WH relative clause*	-2.2	4.1**	-2.5
Phrase-level complexity features			
Premodifying adjective	1.6	0.6	-1.6
Premodifying noun	-0.3	-1.3	1.5
Prepositional phrase <i>of</i> *	-3.5	-3.4	5.6**
Prepositional phrase other than <i>of</i> *	-2.9	0.6	0.7

Note. Single asterisk (*) indicates the complexity features that made great contributions to the association. Double asterisk (**) indicates the main source of complexity at each proficiency level.

Based on the cut-off value $|2.0|$ (Agresti, 2018), a total of five complexity features, marked with single asterisk (*) in the table, have large adjusted residuals, indicating that these five features were significant contributors to the association between L2 writing proficiency and grammatical complexity. By contrast, the remaining four, which failed to produce residuals greater than $|2.0|$, made little contribution to the association.

To be more precise, the large adjusted residuals of the five complexity features

derive from the following eleven cells: (1) finite adverbial clause and the low-rated essays (3.2); (2) finite adverbial clause and the mid-rated essays (2.0); (3) finite adverbial clause and the high-rated essays (-4.0); (4) noun complement clause and the low-rated essays (-2.0); (5) WH relative clause and the low-rated essays (-2.2); (6) WH relative clause and the mid-rated essays (4.1); (7) WH relative clause and the high-rated essays (-2.5); (8) prepositional phrase *of* and the low-rated essays (-3.5); (9) prepositional phrase *of* and the mid-rated essays (-3.4); (10) prepositional phrase *of* and the high-rated essays (5.6); and (11) prepositional phrase other than *of* and the low-rated essays (-2.9). As such, the results of the residual analysis specify which complexity features produced by which proficiency groups made major contributions to the statistical significance of the association.

The positive or negative adjusted residual values are based on comparing observed and expected frequencies (Sharpe, 2015) (See Appendix 3 for further details of the observed and expected count output from SPSS). Positive values indicate that the observed cell frequencies are greater than the estimated expected frequencies and negative values indicate the opposite. According to value types, cells with large adjusted residuals are divided into two categories for each proficiency level:

1. The low-rated essays contained finite adverbial clause more than expected while containing noun complement clause, WH relative clause, prepositional phrase *of* and other than *of* less than expected.
2. The mid-rated essays contained finite adverbial clause and WH relative

clause more than expected while containing prepositional phrase *of* less than expected.

3. The high-rated essays contained prepositional phrase *of* more than expected while containing finite adverbial clause and WH relative clause less than expected.

Especially noteworthy is that the main source of complexity or the largest positive adjusted residual value at each proficiency level, marked with double asterisk (**) in Table 4.3, shows the development that gradually progresses from relying on clausal elaboration to phrasal modification, which is remarkably congruent to the hypothesis suggested by Biber et al. (2011). Specifically, the low-proficiency group showed heavy reliance on earlier acquired clausal features like finite adverbial clause (Stage 2), but used a wide range of phrasal features far less frequently than expected; the intermediate-proficiency group showed a strong preference for later acquired clausal features such as WH relative clause (Stage 3), but still seemed to have difficulties in using phrasal features productively; and the high-proficiency group drew very heavily on prepositional phrase *of*, which is acquired at the higher developmental stages (Stages 3-5), but used basic clausal features, namely, finite adverbial clause and WH relative clause, less than expected.

Put differently, the development of grammatical complexity proceeds from (i) the reliance on finite dependent clauses that function as clause constituents; through (ii) the heavy use of finite clause types that function as NP constituents; to finally, (iii) the marked preference for phrasal structures that function as NP

constituents. Accordingly, as summarized in Table 4.4, the developmental pattern for Korean college students can be accounted for by adopting the two parameters proposed by Biber et al. (2011) as critical determinants of development.

TABLE 4.4
The Observed Developmental Stages for Korean College Students

Proficiency Level	Low	Intermediate	High
The Main Source of Complexity	Finite adverbial clauses	WH relative clauses	Prepositional phrases (of)
Parameter A (Structural type)	Finite dependent clauses	Finite dependent clauses	Dependent phrases
Parameter B (Syntactic function)	Constituents in clauses	Constituents in NPs	Constituents in NPs

Similar developmental patterns can be reasoned in Taguchi et al. (2013), while quite different patterns emerged in Parkinson and Musgrave (2014) and Lan, Lucas, and Sun (2019). Taguchi et al. (2013), as reviewed in Section 2.3.2, reported that subordinating conjunctions and prepositional phrases as postnominal modifiers were distinctive features of the low- and high-rated essays, respectively, as in the current study. With respect to WH relative clauses, no significant difference was detected, which might be due to the comparison of only two groups with high and low proficiency. If that is the case, this result can be interpreted as tentatively suggesting the intermediate level of complexity of WH relative clauses. Consequently, it may be reasonably assumed that Taguchi et al. (2013) showed the developmental trend starting from subordinating conjunctions, through WH

relative clauses, and finally to heavy use of postmodifying prepositional phrases, which closely aligns with the findings of the present study.

Parkinson and Musgrave (2014) and Lan, Lucas, and Sun (2019), who only included nominal modifiers as study variables, stand in stark contrast to Taguchi et al. (2013) and the present study in the following two points. First of all, *of* phrases turned out to be unreliable indicators of proficiency differences in Parkinson and Musgrave (2014). Even close association of *of* phrases with low-proficiency writers was reported in Lan, Lucas, and Sun (2019). Secondly, regarding relative clauses, both studies observed their connection with high-proficiency writers.

In particular, the different developmental patterns shown in Lan, Lucas, and Sun (2019) and the present study can be considered in relation to learners' L1 background because these studies analyzed writing samples produced by L1 Chinese and L1 Korean writers, respectively, unlike other studies using data from different L1 groups. In Chinese language, noun-modifying elements “must occur in front of that noun” (Li & Thompson, 1981, p. 104) and the semantic relationship between two clauses is not explicitly stated but “must be inferred” (ibid., p. 641). Considering these characteristics, one could postulate that L1 Chinese writers can find post-noun-modifying clausal features of English, such as relative clauses, most challenging. On the other hand, most *of* phrases used by low-proficiency students in Lan, Lucas, and Sun (2019) can be considered easily accessible to

L1 Chinese writers because they were not true postmodifiers of a noun, occurring after package nouns (e.g., *kind of*). Korean also has a head-final NP structure but relative flexibility in combining clauses (Kang, 2005). These characteristics support the relative ease of acquiring clausal modifiers, such as relative clauses, compared to phrasal modifiers, such as prepositional phrases. Besides, the requirement that a modifying clause must precede a head noun in Korean language prevents the use of multiple modifying clauses due to parsing difficulties; by the effect of that, “multiple short sentences combined together with conjunctions” are preferred in Korean (Baik, 1994, p. 162). It thus seems reasonable to hypothesize that finite adverbial clauses would be readily acceptable to L1 Korean writers, as shown in the result of the present study. The causal links between L2 features preferred at each proficiency level and L1 characteristics may offer evidence for possible L1 impacts on the use of complexity features in L2 writing.

In fact, as will be discussed in the following sections, it may be insufficient and unreasonable to predict or describe developmental patterns based purely on frequency information about grammatical features. Thus, these frequency-based quantitative findings will be supplemented by qualitatively reviewing the actual uses of complexity features strongly associated with L2 writing proficiency in Section 4.4 and only tenuously related features in Section 4.5.

4.4 The Grammatical Complexity Features with Great Contribution to the Association

The results of residual analysis revealed that the robust association between grammatical complexity and L2 writing proficiency was derived primarily from three clause-level complexity features (i.e., finite adverbial clause, noun complement clause, WH relative clause) and two phrase-level complexity features (i.e., prepositional phrase *of* and other than *of*). In particular, it was noteworthy that finite adverbial clauses turned out to be the most significant characteristic of the low-rated essays; WH relative clauses were the most predominant feature of the mid-rated essays; and *of* phrases proved to be the strongest indicator to distinguish the high-rated essays from the low- and mid-rated ones. The present section elaborates on the differences among the three proficiency groups in the use of these complexity features by qualitatively evaluating student writing from the perspective of lexical diversity, accuracy, and complexities typical of academic written registers.

4.4.1 Finite Adverbial Clauses

The residual analysis suggests that the principal complexity feature distinguishing proficiency groups at the clause level is finite adverbial clauses (*because, if, although*), as they are clausal features with the largest sum of absolute

values of adjusted standardized residuals. Specifically, finite adverbial clauses were produced far more than expected in the low-rated essays (3.2) and the mid-rated essays (2.0) but much less than expected in the high-rated essays (-4.0). For example, Excerpt 1 (a) from a low-rated essay includes four adverbial clauses (one *because* clause, three *if* clauses) in the 129-word writing. Besides, the writer produced a series of the coordinator *but*. Previous corpus-based studies showed that the coordinator *but* is considerably more common in conversational discourse than in formal writing, unlike the coordinator *and*, which prevails in written texts as a phrasal connector (Biber et al., 1999).²⁰ Therefore, the extensive use of both clausal subordination and coordination in the low-rated essay can be viewed as indicative of the less advanced students' undue reliance on complexity features characteristic of conversation in formal writing. Similarly, Excerpt 1 (b) from a mid-rated essay showed the frequent occurrences of adverbial clauses (four *if* clauses, four *because* clauses) and the coordinator *but* in the 280-word writing.

Excerpt 1 (a) <Low-rated essay, file no. 2631>²¹

I think, **if** animals was killed **but** humans will be alive, it is sadly not bad. **but** non-mean killing animals is very crucial. **because** i always think the animal's mean is contain the human. killing animal same merder. Medical experiments can help human? i say no, it just kill animal and kill human. **if** human live alone, it maybe yes. **but** human cannot live alone. it need

²⁰ Longman Grammar of Spoken and Written English (Biber et al., 1999) is grounded upon comparing four registers: conversation, news, fiction, and academic prose. It is thus often cited in this study as an ideal resource conducive to identifying (dis)similarities between student writing and academic prose and observing whether student writers are progressing towards the target register.

²¹ The spelling errors in the sampled essays were not corrected.

the animals that human live. Medical experiments will alive the several human, **but** same valued animals will die. The world have to keep valence. **if the world cannot keep valence**, every animals, humans, plants will be die by world to keep the valence. So human find other way to develop the culture. that's a real work scientist have to.

Excerpt 1 (b) <Mid-rated essay, file no. 1913>

I think it should be allowed. Without the punishment, students would do anything they want! **But** also it should not be severe. **if the punishment is severe**, punished students becomes angry and do bad to the punishing teacher. That is, the punishment should permitted appropriately. I read some articles talking that some teachers bullying their students. There is struggle insisting not allowing the punishment **because this idiot teachers**. I think everything in the world should be appropriate. For example, **if love is much bigger than the appropriate state**, someone would be hurted **because he or she should let her or him go because the other someone could do better than he or she does**. Isn't it a bad situation? I wrote the punishment is needed to be allowed. **But if a teacher is a good teacher**, it would not be needed. So, I want to talk all of teachers that, "You listen to your students." The students does bad **because the things are not going to be that they want to be**. You teachers should understand them in your hearts, listen to their hurted minds **if you want to be a good teacher and don't want to punish your students**. [...]

Not only is excessive subordination unsuited for academic written register, but the immoderate subordination in writing, such as “a chain of *if* subordinate clauses” in Excerpts 1 (a) and (b), is problematic as it may lead to “an overly complex sentence that is difficult to follow” (Taguchi et al., 2013, p. 426). Thus, more advanced writers might have expressed the same content by means of phrasal modification instead of subordination for clear and condensed information. To illustrate, the reduction of the first *if* subordinate clause in Excerpt 1 (b) (*If the punishment is severe*) into an NP with an attributive adjective (e.g., *severe*

punishment) that could function as the subject of an independent clause, may provide the same information in a compact and easy-to-understand form. Examples (4) and (5) below illustrate again this point; the *if* clause is in bold; attributive adjectives are in italics.

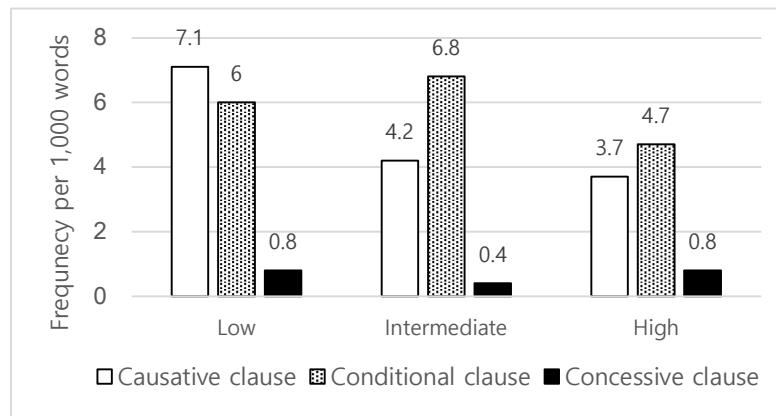
- (4) **If the punishment is severe**, punished students becomes angry and do bad to the punishing teacher. <Mid, file 1913>
- (5) Surely, *excessive physical punishment* can be a problem. <High, file 1061>

In fact, finite adverbial clauses were not infrequent in the high-rated essays in absolute terms, but their density turned out much lower than that of the low- and mid-rated essays, given the normed frequencies of finite adverbial clauses (see Table 4.1). The low- and mid-rated essays are thus likely to be judged more structurally complex than the high-rated essays when measuring complexity solely via finite adverbial clauses, disregarding the different kinds of complexity suited for academic writing (Biber et al., 2011; Halliday, 1989, 2004; Rimmer, 2006).

Surprisingly, though, when looking at causative, conditional, and concessive adverbial clauses separately (see Figure 4.3), their distributions followed somewhat different patterns according to proficiency levels. More specifically, in contrast to causative clauses, which showed the salient pattern of decreasing in frequency towards the higher proficiency levels, conditional and concessive clauses showed no such pattern, partially contradicting Biber et al.'s (2011) developmental index. The three types of adverbial clauses showed a significant

association with the three proficiency levels ($\chi^2= 13.053$, $df= 4$, $p= .011$).

FIGURE 4.3 Frequencies of Finite Adverbial Clauses



The findings of this study concerning conditional and concessive clauses are in line with Staples and Reppen (2016), who reported that English L1 writers, who can be thought of as more advanced writers, make greater use of these two types of adverbial clauses than L2 writers, as opposed to the case of causative clauses. Biber et al.'s (1999) large-scale corpus analyses further support these findings, showing that conditional clauses, though the majority is found in conversational discourse, are fairly common in written discourse, and concessive clauses are marginally more frequent in written texts than in conversational discourse. Empirical evidence that comes from Biber et al. (1999), Staples and Reppen (2016), and the current study consistently upholds the need to re-examine Biber et al.'s (2011) developmental index that categorized all adverbial clauses into the same stage without considering "lexical choices among subordinators" and

associated functional differences (Staples & Reppen, 2016, p. 21).

The status of conditional and concessive clauses in academic written discourse as “important contributors to the development of arguments, which is a significant goal of academic writing” (Biber et al., 1999, p. 825) may come from their role as hedges (Hinkel, 2013; Hyland, 1994; Warchała, 2010). The hedging function of *if* clauses is to limit the certainty of an argument by conditioning its validity to other factors (Warchała, 2010). This is well illustrated in Examples (6) through (9), where *if* clauses were used to specify certain conditions under which the claims hold true. The correlative *then* in the main clauses serves to clarify the connection between the two statements (Biber et al., 1999), but no such case was found in the low-rated essays. Concession clauses can also be seen as “sophisticated hedging devices” in academic discourse, allowing the writer to express a balanced view of the subject matter (Hinkel, 2013, p.11). The hedging function of concessive clauses is manifest in Examples (10) and (11), where *although* clauses were used to indicate the limitations of specific arguments or facts.

- (6) **If using cellular phones is allowed, *then* watching movies using navigation utilities should be allowed to.** <Mid, file 2810>
- (7) **If someone want to smoke in public buildings, *then* he/she should try hard not to harm other people around him/her.** <High, file 478>
- (8) **If that two rights are conflicted, *then* I think smokers should understand non-smokers.** <Mid, file 492>
- (9) **If we have any back up plans for the chaos *then* it's okay to just ban physical punishment because we may still have some way to control students who behave bad.** <High, file 1746>
- (10) But **although smokers harm non-smoking people**, they have right to

smoke personally. <Mid, file 943>

- (11) **Although** the development of army weapons has been achieved, still the number of military army takes great charge of the military forces. <High, file 2761>

In light of the use of conditional clause as formal hedges in academic discourse, Warchał (2010) argued that they serve “interpersonal functions” aimed at “creating conditions in which the author’s claim can be favorably received by the reader” (p. 149). He further noted that *if* clauses may enhance their “consensus-building potential” by agreeing with other interpersonal features like modal auxiliary verbs (e.g., *might*, *could*) or adverbial modal expressions (e.g., *perhaps*, *possibly*) (p. 149). This may also be the case for concessive clauses that have similar functions. In this sense, the role of conditional and concessive clauses as an effective tool for creating consensus in academic discourse could not be completely replaced by condensed phrasal structures.

Although as illustrated in Excerpts 1 (a) and (b), *if* clauses were frequently used in some lower-rated essays, they rarely performed their prototypical function in academic writing. Instead, most of them seemed to be related to their conversational function that introduces “problems as hypothetical” when making negative comments (Ford, 1997, p. 401). In other words, *if* clauses in the lower-rated essays served primarily to indicate that the problems that the writer delivered in *if* clauses are “provisional, not strongly claimed to be true” (Ford, 1997, p. 393) as in Example (12) from Excerpt 1 (a). Another conversational function of *if*

clauses that “mitigates the force of the suggestion” (Biber et al., 1999, p. 821; Biber, 2006; Ford, 1997) is well demonstrated in Example (13) from Excerpt 1 (b). The *if* conditional in this sentence indicates that a reader has the choice to decide whether to accept the author’s argument presented in the main clause, as if conditionals in interactional conversation cushion the intensity of directives or suggestions by leaving them as a hearer’s choice.

- (12) **If the world cannot keep valence**, every animals, humans, plants will be die by world to keep the valance. <Low, file 2631>
- (13) You teachers should understand them in your hearts, listen to their hurted minds **if you want to be a good teacher and don't want to punish your students**. <Mid, file 1913>

Taken together, finite adverbial clauses were generally much more associated with the low- and mid-rated essays than the high-rated ones. However, an in-depth review of the writing samples suggested that conditional and concessive clauses are distinct from causative clauses, given their prominent role as hedges that help the development of the author’s argument. It may be for this reason that the occurrences of these two adverbial clauses in the high-rated essays are not less than those in the low- and mid-rated essays. Thus, although subordination is generally a characteristic of conversational language rather than formal writing, the proper use of conditional and concessive clauses as hedges should be encouraged for successful academic writing in establishing interpersonal relations, while correcting the excessive use of adverbial clauses that obscures the argument

by promoting subclausal elaboration instead.

4.4.2 Prepositional Phrases as Nominal Postmodifiers

Post-noun-modifying *of* phrase was the second most common type of grammatical complexity feature after attributive adjective (see Table 4.1). This is attributed to “a range of uses in expressing a close semantic relationship between the head noun and the following noun phrase” (Biber et al., 1999, p. 636).

Considering that *of* phrases made up a large proportion of complexity features in student texts, the residual analysis results showing notable differences in their frequencies among the three proficiency groups are quite surprising. Specifically, *of* phrases were much more frequent than expected in the high-rated essays (5.6) while far rarer than expected in the mid-rated essays (-3.4) and low-rated essays (-3.5). As a result, prepositional phrase *of* was the grammatical feature with the largest sum of absolute values of adjusted residuals among all the complexity features analyzed, indicating that *of* phrases contributed most to the association between the use of complexity features and writing proficiency levels. In other words, they proved to be the best discriminator of proficiency-level differences. Regarding prepositional phrases other than *of* (*in, on with, for*), the low-proficiency group (-2.9) used much less than expected as in the case of prepositional phrases *of*, but the high-proficiency group (0.7) and the mid-proficiency group (0.6) used moderately more than expected. The overall results

are in alignment with the findings of Kim (2020) and Taguchi et al. (2013) that prepositional phrases, particularly *of* phrases, are positively associated with L2 writing proficiency.

Excerpt 2 (a) from a high-rated essay demonstrates the dense use of postmodifying prepositional phrases in advanced writing. In the 299-word essay, fifteen prepositional phrases beginning with the prepositions *of*, *for*, and *with* occurred to modify the preceding head NPs. A great majority of them are *of* phrases, as Biber et al. (1999) noted that the preposition *of* made up 60-65% of all postmodifying prepositional phrases across registers. Excerpt 2 (a) also illustrates how other phrasal modifiers are embedded in prepositional phrases to build more compressed structures in advanced writing. To be precise, prepositional phrases in the following excerpt include attributive adjective (i.e., *medical*, *infectious*), participial adjective (i.e., *tested*), or premodifying noun (i.e., *drug*), resulting in greater lexical density and grammatical complexity; subordinators and controlling verbs are in bold with the clauses they introduce underlined; prepositional phrases are bold underlined with the head NPs in brackets; premodifying adjectives are in italics; premodifying nouns are in bold italics.

Excerpt 2 (a) <High-rated essay, file no. 2546>

[Using animals] **for the purpose of testing drugs** is justifiable and necessary: animals do not have rights equal to [those] **of humans**, and even if they did, [using them] **for medical experiments** must be done for the *greater* good. Some people **argue that animals have rights too**, and **that we are not entitled to infringe on them**. However, their rights are non-

existent or negligible compared to [those] **of humans**. Humans have more rights **because they live for higher causes**. We have more *advanced* needs than just to feed or reproduce. We go to museums, we create works of art, and more importantly, we know how to love each other spiritually. [The size] **of rights** is determined by [the kinds] **of causes** people or animals live for. Thus, animals have almost no rights compared to us. But even if animals did have *equal* rights to humans, we should still be allowed to experiment with them for the *greater* good. **If we were to let drugs hit the market without being thoroughly tested**, we would face *dreadful* consequences: *child* deformation, [spread] **of infectious diseases**, and even [extinction] **of species**. Among them, [the outbreak] **of diseases**, [the most *likely* result] **of the production of lightly tested drugs**, would not just affect humans. It would spread out to the *entire* ecosystem, bringing *detrimental* harms to animals as well. To prevent such *irreparable* damages, we should take as *through* steps as possible in testing drugs. [Experimenting] **with animals** is the *best* way to ensure [accuracy] **of the process**. In a nutshell, animals do not have *comparable* rights to *human* beings, and even if they did, some sacrifices should be made to prevent *horrific* consequences. For these reasons, [using animals] **for drug experiments** is not only the *right* thing to do, but also necessary.

Excerpt 2 (b) from a low-rated essay further supports the notion that prepositional phrases as nominal postmodifiers are significant proficiency predictors. This essay includes no post-noun-modifying prepositional phrase, showing heavy reliance on clause-level complexity features, namely, finite adverbial clauses and verb complement clauses typical of conversational discourse. In this relatively short 172-word essay, *because* and *if* clauses appeared four times in total, and *that* complements controlled by the verb *think*, which is the particularly frequent controlling verb in conversation (Biber et al., 1999), appeared three times.

Excerpt 2 (b) <Low-rated essay, file no. 829>

I **think that** physical punishment should be allowed in school. (but it is only OK when students have *bed* attituded to their teacher.) thesedays, many student don't listen advise. **because** they search the internet about their problem not teacher but computer. then they were seem to trust internet data then teacher's advise. so they would reject teachers *mental* punishment. and their respond is *dangerous* level. For example, In 2010, Seoul **education** department start teacher to no hit and they give penalty to student like 'Don't go to school. 2 days!' maybe, they **think that** punished-student would be ashamed because they can't go to school. but punished-student dont' think it. they boast their penalty. **because they do not need to go school and they enjoy playing the game that times. And forethefore other students envy penalty. In this situation, I **think that** students should be hitted. *generous* people dislike *physical* punishment. students are also same. **If** teachers treated their student like animal in punishment, student would listed teachers advise.**

Excerpt 2 (b) from a low-rated essay is in stark contrast with Excerpt 2 (a) from a high-rated essay containing only a few adverbial clauses (one *if* clause, one *because* clause) and one complement clause with the controlling verb *argue*, despite its almost twice the length of the low-rated essay. The comparison of the two essays clearly demonstrates a decrease in clause-level complexity but an increase in phrase-level complexity, especially via *of* phrases, towards the upper proficiency level. This suggests that decreased complexity at the clausal level may indicate a positive change in L2 writing development.

Another interesting finding is that the more proficient students drew more on multiple prepositional phrases than the less proficient students, though not by a

massive margin. For example, in Excerpt 2 (a) above, the postmodifier complex²² following the two head NPs (i.e., *using animal, the most likely result*) contains two prepositional phrases (i.e., *for the purpose of testing drugs, of the production of lightly tested drugs*). Overall, multiple prepositional phrases as nominal postmodifiers were used most in the high-rated essays (1.6 instances per 1,000 words), and no difference was found between the low- and mid-rated essays (0.8 instances per 1,000 words in both). As shown in Examples (14) through (16), the first in a sequence was *of* phrase, and the second modifier was *in* phrase in approximately half of all the multiple prepositional phrases in student writing. Interestingly, in the case of the high-rated essays, far more complex and compressed NPs including three prepositional phrases emerged, while no such postmodification occurred in the mid- or low-rated essays. The use of three prepositional phrases as postmodifiers was not frequent in the high-rated essays as well, showing only two occurrences as in Examples (17) and (18). In both cases, the first two postmodifiers in a sequence were *of* phrases with the third being *in* phrase in the same manner as the common combinations of two prepositional phrases.

(14) The physical punishment **of children in schools**. <Low, file 547>

(15) My idea **of banning smoking in public buildings**. <Mid, file 2181>

(16) More loss **of invaluable lives in traffic accidents**. <High, file 2552>

(17) The number **of cases of physical punishment in schools**. <High, file 495>

(18) The abolition **of the physical punishment of children in schools**. <High,

²² A postmodifier complex refers to “a combination of postmodifiers” (Biber et al., 1999, p. 641).

file 1437>

These findings are consistent with previous studies showing that multiple prepositional phrases are positively related to writing proficiency levels (Kim, 2020) and that NPs with multiple postmodifiers, particularly multiple prepositional phrases, represent advanced formal writing (Biber, 2006; Biber et al., 1999; Biber et al., 2011). Accordingly, NPs with multiple prepositional phrases are placed in Stage 5 of Biber et al.'s (2011) framework, while those with a single prepositional phrase are in Stages 3 and 4, indicating that the productive use of multiple prepositional phrases is “one of the main linguistic challenges that students encounter as they progress through a university education and learn to deal with written academic registers” (Biber, 2006, p. 76). Given that the essays examined in the present study were produced by first-year college students who had just started university courses, it is not surprising that multiple prepositional phrases exhibited no marked difference among the proficiency groups, and that even the high-proficiency group did not appear to have reached the developmental level to use such structures productively.

Regarding the use of package nouns followed by *of* phrases, the findings of this study seem inconsistent with Lan, Lucas, and Sun (2019), in which low-proficiency students drew heavily on *of* phrases to modify quantifying nouns (e.g., *number of*, *lot of*), partitive nouns (e.g., *part of*), and species nouns (e.g., *kind of*). The current study, by contrast, showed that the high-rated essays had a slightly

higher frequency of package nouns with postmodifying *of* phrases (2.4 instances per 1,000 words), followed by the low-rated essays (1.3 per 1,000 words) and the mid-rated essays (1.1 per 1,000 words). Despite no significant difference in frequency among the proficiency groups, a qualitative check revealed that the students with higher proficiency used a much wider range of package nouns (e.g., *lot of, number of, amount of, thousands of, millions of, years of, deal of, level of, handful of, degree of, kind of, type of, bunch of, group of, piece of, series of*), whereas the low-rated essays showed the repetitive use of only three types of package nouns (i.e., *lot of, kind of, hundreds of*). Besides, the higher-proficiency students made them more complex with the addition of premodifying adjectives (e.g., *growing, enormous, uncountable, specific*) as in Examples (19) through (23), whereas no such cases were found in the low-rated essays.

- (19) These days, a *growing number of people* usually think that smoking in public buildings is acceptable by individuals. <Mid, file 938>
- (20) Also, experimenting new drugs on them can save an *enormous number of people or animals*. <Mid, file 2294>
- (21) However, I saw a number of teachers who discipline their students well with no rods and also an *uncountable number of students* who obey their teachers without being hit. <High, file 2872>
- (22) Teachers are incapable of inducing an *exact amount of physical punishment* and thus often go overboard. <High, file 1199>
- (23) Moreover, harshness of teacher to *specific type of students* will be reduced. <High, file 1757>

Taken together, although package nouns followed by *of* phrases themselves might be associated with simple grammatical functions (Lan, Lucas, & Sun, 2019),

it turned out that the higher-proficiency students tended to combine them with other phrasal modifiers, achieving greater complexity and carrying more accurate information. Besides, as indicated in the extraction criteria for package nouns in this study (see Section 3.3), certain types of species nouns (e.g., *type(s) of*, *species of*) are more prevalent in written discourse than in conversation due to their classifying function required in academic writing (Biber et al., 1999). This suggests that some package nouns should be treated as essential features of academic writing rather than dismissing them as features serving simple grammatical functions.

In summary, prepositional phrases *of* were found to be the grammatical complexity features that best represented the characteristics of the high-rated essays, considering their largest positive residuals associated with this structure. In addition, among all the complexity features, *of* phrases turned out to be the most reliable predictor of L2 writing proficiency, showing the biggest difference in adjusted standardized residuals between proficiency levels. Besides the high frequency of postmodifying *of* phrases, the qualitative analysis further revealed that more phrasal embedding in prepositional phrases and the emergence of multiple prepositional phrases best mirrored the developmental progression of L2 writers from clausal to phrasal complexity.

4.4.3 WH Relative Clauses

WH relative clauses (*who, which, whom, whose*) were used much more than expected in the mid-rated essays (4.1) but much less than expected in the high-rated essays (-2.5) and the low-rated essays (-2.2). Thus, these structures are clearly distinguished from other complexity features in that only the mid-proficiency group used significantly more than expected. This result seems to be associated with the intermediate nature of complexity concerning WH relative clauses. Considering the two parameters which determine the degree of complexity (see discussion in Section 2.2.2), relative clauses (*that* and WH) are “mixed or intermediate on the two parameters” as “finite clause types functioning as a constituent in a noun phrase” (Biber et al., 2011, p. 27). These intermediate features generally exhibit no significant difference in their frequencies between formal written and spoken registers as in the case of *that* relative clause, but WH relative clause is exceptional as being much more common in written texts than in conversation. Thus, increasing use of WH relative clauses might reflect a relatively advanced stage of development than the stage of using other types of clausal features such as finite adverbial clauses (Biber et al., 2011).

However, the results of prior L2 writing studies on whether relative clauses (*that* and WH) are indicative of writing development are often mixed and difficult to aggregate (Kang & Oh, 2022; Lan et al., 2022; Lan, Lucas, & Sun., 2019; Lan & Sun, 2019; Parkinson & Musgrave, 2014; Taguchi et al., 2013). The discrepancy

might result from level classification. None of the studies examined in this paper involved the comparisons of multiple levels of proficiency when examining relative clauses in relation to L2 writing proficiency. Instead, L2 writers fell largely into two levels of proficiency (i.e., high and low). In such research design, grammatical structures achieved at intermediate levels of development can be confused with or disguised as ones achieved at other developmental stages. Another reason for the inconsistent results on the significance of relative clauses as indicators of L2 writing proficiency might be the combined analysis of *that* relatives and WH relatives in most studies despite their different distributional patterns across registers.

The result of the present study based on the multiple-level comparisons and separate analysis of these two types of relative clauses clearly supports the intermediate stage of WH relative clauses (Biber et al., 2011), showing their large positive adjusted residuals associated with the mid-level students (4.1). Besides, the large negative adjusted residuals associated with the low-level students (-2.2) suggest that WH relative clauses might not yet be readily available at their proficiency level, whereas the large negative residuals associated with the high-level students (-2.5) indicate that they might have already passed the intermediate stage, favoring phrasal elaboration over clausal modifiers. In contrast, the absolute values of *that* relative clauses are lower than the cut-off value |2.0| at all proficiency levels, which also bears out the previous finding of register research

that *that* relatives are equally common in all registers with no salient relationship with written discourse (Biber et al., 1999; Biber et al., 2011). The relatively large positive adjusted residual value associated with the low-rated essays (1.7), though slightly lower than the cut-off value, suggests that *that* relatives were quite common in the low-rated essays and by far preferred over WH relatives (-2.2) as in Taguchi et al. (2013), probably because the relative pronoun *that* is available without worrying about “+/- human feature of the NP for which it substitutes” (Larsen-Freeman & Celce-Murcia, 2016, p. 618).

Further evidence for the intermediate stage of WH relative clauses comes from the comparison of the accuracy with which WH relative clauses were used. The error rates of the low- and high-proficiency students (5.3% and 8.7%, respectively) were considerably lower than that of the mid-proficiency students (23.6%). The low error rate for these structures in the low-proficiency group could be explained by avoidance strategies for error-prone structures (Brown, 2000). The absolute frequency of WH relatives in the low-rated essays was only less than one-fifth of that in either the mid- or the high-proficiency group (see Table 4.1), which possibly reflects that the low-proficiency students avoided the use of WH relative clauses (Stage 3) rather than taking the risk of making mistakes and may, for instance, use the complexity features in the basic stage instead (Stages 1 and 2). In contrast, the low error rate of the high-proficiency students, despite being with a high occurrence, might indicate that they were already beyond the developmental stage

where WH relative clauses are acquired as noun modifiers. The majority of errors across proficiency levels related to subject-verb agreement, that is, agreement between the head NP being modified and the verb in a relative clause, while some involved the wrong choice of relativizers, as shown in Examples (24) through (26). These findings are in line with Kang and Oh (2022), who identified number agreement as the major source of errors in relative clauses found in YELC 2011. They regarded this type of error as an “L1-induced error” given that Korean has no grammatical feature that serves the function of subject-verb agreement (p. 10).

- (24) All thing **who** move and think is same lives. <Low, file 757>
- (25) I think that those who have religion **which** don't let them use military force should be reconsidered and excluded from military observation. <Mid, file 3077>
- (26) Teachers shouldn't be ones **whom** try to physically punish a student, but be the ones **whom** respect students' feelings and their backgrounds. <High, file 1741>

In addition, although the relative clauses in Examples (27) and (28) are not grammatically problematic, it is reasonable to hypothesize that instead of using relative clauses, advanced writers might have used *of* phrases (i.e., *freedom of smokers*) and attributive adjectives (i.e., *other innocent people*), respectively, considering that phrasal embedding is preferred in academic writing over clausal elaboration for a high degree of lexical density and greater conciseness (Taguchi et al., 2013).

- (27) Smoking is freedom **that smokers have**. <Low, file 1925>
(28) And car accident can hurt other people **who is innocent**. <Mid, file 3077>

A close investigation into the choice of relativizer made by each proficiency group revealed that besides showing the overall low frequency of WH relatives, the low-rated essays only involved the relativizer *who* with no instances of the remaining three (i.e., *which*, *whose*, *whom*). Similarly, *who* was by far the most frequent in both the mid-rated essays (72.4%) and the high-rated essays (82.5%); however, unlike in the low-rated essays, *which* was also relatively common in the mid-rated essays (23.6%) and the high-rated essays (10.7%), though *whom* and *whose* were still rare in both sets of writings. Such findings may be in part due to the small-sized corpus examined in this study but adequately suggest the lack of ability to deal with “NPs in a range of syntactic positions” by using “object pronoun *whom* or possessive determiner *whose*” (Larsen-Freeman & Celce-Murcia, 2016, p. 618).

While the results generally aligned with the previous finding that advanced writers produced relative clauses with various types of relativizers (Lan, Lucas, & Sun, 2019; Lan & Sun, 2019), the distribution of relativizers in the Korean learner corpus was distinctive in that *who* was the most common relativizer at all proficiency levels. This result contrasts with the finding of Lan and Sun (2019) that L2 writers’ use of relative clauses was restricted to *which* at initial stages of development and showed increasing diversity as writing proficiency improved.

Register research also reported that the most common relativizer in formal writing is *which* as “there are relatively few animate references at all” (Biber et al., 1999, p. 612). Thus, *who*, which usually occurs with animate head nouns, is bound to be far less frequent than *which* in academic written register. The strikingly different distributions of *which* and *who* in the Korean learner corpus might reflect how writings of Korean college students at all proficiency levels are strongly associated with the personal style of conversational discourse, diverging from academic writing style, but on the other hand, this can be attributed to topic effects on the choice of relative pronouns. The six topics given in argumentative essays of YELC 2011 are all relevant to human life or individual well-being (e.g., smoking in public places, physical punishment at schools, mandatory military service), which are likely to induce student writers to argue based on their daily life experiences. Thus, such human life-related topics may have created more animate references that required the use of *who* compared to typical academic texts such as research articles analyzed in Biber et al. (1999).

Although most WH relative clauses in student writing were introduced by *who*, it is relatively often shown in the mid-rated essays that relative clauses beginning with different relativizers are embedded in a single sentence, as illustrated in Examples (29) through (31). This was true for 11.4% of the occurrences of WH relative clauses in the mid-rated essays. However, such cases were not as often in the high-rated essays (3.9%) and non-existent in the low-rated essays, which

suggests again that WH relatives were the most preferred means at intermediate levels to achieve structurally compressed NPs.

- (29) Also some men **whose** family is so poverty-stricken and there is no person **who can earn money** if he go to army should ne extracted from military service obligation list. <Mid, file 1111>
- (30) I think that those **who have religion** **which** don't let them use military force would be reconsidered and excluded from. <Mid, file 3077>
- (31) The government should build more counseling facillities **which can give practical help to the students** **who are in conflict with their teachers.** <High, file 2374>

To sum up, WH relative clauses proved especially characteristic of intermediate developmental stages of L2 writing, considering their largest positive residuals associated with the mid-rated essays. This observation was further supported by the low-level writers' avoidance of these structures and the high-level writers' very few errors involved in using them. The two representative error types concerning the use of WH relative clauses identified the areas for language-focused instruction. The qualitative analysis also demonstrated the need to refrain from using the relativizer *who* typical of conversational discourse, and to use phrasal modifiers in place of relative clauses when phrasal embedding can increase the economy and conciseness of information delivery.

4.4.4 Finite Complement Clauses Controlled by Nouns

The large negative adjusted residual value for noun complement clauses

associated with the low-proficiency group (-2.0) indicates that there were much fewer noun complement clauses in the low-rated essays than would be expected. In contrast, the positive adjusted residual values associated with the high-proficiency group (1.1) and the intermediate-proficiency group (0.2), though they failed to exceed the designated cutoff value $|2.0|$, indicate that there were moderately more noun complement clauses in their essays than would be expected. These results appear to contradict the general developmental progression from clausal to subclausal complexity. However, as shown in Table 2.2, noun complement clauses are placed in Stage 5 of Biber et al.'s (2011) framework, whereas those with similar structures, namely, complement clauses after verbs or adjectives, are placed in Stages 1-3 of the model. The different complexities of these seemingly similar structures are based on Biber et al.'s (2011) corpus finding that "parameter B (syntactic function) continues to be important, even for the structural categories of finite and nonfinite dependent clause" (p. 28). In other words, whether the structure functions as a constituent in NPs is prioritized over structural type in the assessment of its complexity. This is particularly the case for noun complement clauses, given their prevalence in academic writing but a rarity in conversation, and the exact opposite pattern for complement clauses headed by verbs or adjectives (Biber et al., 1999).

The contrasting distributional patterns might have to do with their different functions as stance markers. The stance conveyed by nouns is "not normally

attributed to anyone,” whereas the stance reported by verbs is “directly attributed to participants” (Biber et al., 1999, p. 650). Thus, academic writing, whose primary concern is the information being delivered and not the personal feelings, shows a preference for noun complement clauses, which thus become a useful index of complexification at advanced levels, unlike other complement clauses. Prior studies that explored stance marker use of L2 learners in academic prose have also provided extensive empirical evidence to suggest that the ability to mark stance correlates highly with L2 writing proficiency with more proficient writers using a rich array of stance features (Hu & Li, 2015; Kim & Suh, 2014; Oh & Kang, 2013). The finding of the current study that noun complement structures are the areas where the mid-proficiency group (0.2) and the high-proficiency group (1.1) have a small positive adjusted value while the low-proficiency group (-2.0) has a large negative adjusted value seems to further support the previous finding that these structures are typical of professional academic writing and acquired at the most advanced developmental stage of L2 writing.

Not only did the students at all proficiency levels fail to make productive use of noun complement structures but also showed heavy dependence on a few controlling nouns. Table 4.5 draws the comparison of the normed frequencies of the top six most frequent controlling nouns taking *that*-clauses in the entire sets of writing. The one most frequent overall was *fact* (14.4%). When looking at each proficiency level separately, the most employed controlling nouns in the high-rated

essays were *fact* and *idea* (16.0% each), followed by *opinion* and *problem* (4.0% each). Similarly, *fact* (14.3%) was particularly featured in the mid-rated essays, followed by *opinion*, *reason*, and *news* (8.6% each). The overall frequency of noun plus *that*-clause in the low-rated essays was very low (only 0.8 instances per 1,000 words), with no head nouns occurring more than once.

TABLE 4.5
Top 6 Nouns Controlling *That*-clauses

	Low	%	Intermediate	%	High	%
<i>Fact</i>	0	0.0	25.6	14.3	33.5	16.0
<i>Idea</i>	0	0.0	5.1	2.9	33.5	16.0
<i>Opinion</i>	0	0.0	15.3	8.6	8.4	4.0
<i>Reason</i>	0	0.0	15.3	8.6	4.2	2.0
<i>News</i>	0	0.0	15.3	8.6	4.2	2.0
<i>Problem</i>	15.8	20.0	5.1	2.9	8.4	4.0

Note. Normed per 100,000 words.

More important, the low-rated essays were clearly distinguished from the other two sets of writing in the choice of head nouns. In the L2 writing domain, nouns taking *that*-clauses have been highlighted as cohesive devices for achieving textual coherence under different labels such as shell nouns²³ and signalling nouns²⁴ (e.g., Aktas & Cortes, 2008; Flowerdew, 2006; Jang & Rhee, 2014; Oh, 2014). Especially in the context of argumentative writing, these nouns can double

²³ Shell nouns are “nouns which require lexicalisation in their immediate context” (Hunston & Francis, 1999, p. 185).

²⁴ Signalling nouns refer to “potentially any abstract noun, the meaning of which can only be made specific by reference to its context” (Flowerdew, 2003, p. 329).

as persuasive devices by reflecting the writer's perspective on the content presented in the following *that*-clauses. As Wingate (2012) noted, "the analysis and evaluation of content knowledge" could be a basis for "developing an argument" (p. 146). The stance nouns of a writer's choice could in turn affect "readers' comprehension and interpretation of the complement information" thereby making them hold a particular view on the issue (Jiang, 2015, p. 92). In this regard, it is important to note that none of the head nouns found in the low-rated essays (e.g., *memory*, *scene*, *opportunity*) overlapped with stance nouns commonly used in academic prose such as *fact*, *possibility*, *idea*, and *opinion* (Biber et al., 1999; Larsen-Freeman & Celce-Murcia, 2016), while the head nouns found in the mid- or high-rated essays quite corresponded to them. This alignment in the choice of head nouns suggests that the students at upper-intermediate levels started to acquire some generic conventions of academic writing on the use of noun plus *that*-clause.

More in-depth observation of head nouns in student writing was made based on the findings reported by Jiang (2015). Focusing on a particular kind of academic genre, namely, argumentative essays, he found that in contrast to the common stance nouns in L1 writers' argumentative essays (e.g., *argument*, *claim*, *debate*, *opposition*), the common nouns in L2 writers' essays (e.g., *idea*, *opinion*, *view*) are "pragmatically vague and void of illocutionary force" making it difficult to "formulate argumentation and extend the discussions and arguments further" (pp. 96-97). Staples and Reppen (2016) lend further credence to his finding,

showing that the noun *argument* was used frequently by L1 English writers to control complement clauses in argumentative writing, but this noun was never used by the two L2 groups (i.e., L1 Chinese and L1 Arabic). A similar pattern was found in this study based on the Korean learner corpus, given that *fact*, *idea*, and *opinion* were the most frequent head nouns, and there were only two head nouns in the entire dataset that express specific speech acts (i.e., *argument*, *claim*) with very low frequency (0.1 instances per 1,000 in the high-rated essays and no instances in the mid- and low-rated ones). Examples (32) through (35) below demonstrate how the choice of head noun influences the development of the writer's argument.

- (32) There are campaigns not to use animals in medical experimenting and **opinions** that it is necessary in our life. <Mid, file 2294>
- (33) However, I think it is a terrible **idea** that people must use their real name on the Internet. <High, file 2178>
- (34) There is also an **argument** that physical punishment is the best way for the teacher to maintain respect towards the teacher. However, what is mistaken in this **arguement** that teachers and students are partners, not enemies. <High, file 2225>
- (35) In addition, **claims** that animals are especially suffering in medical experiments are exaggerated. Actually our daily lives are able on the basis of animals' sacrifices. <High, file 2516>

Taken together, the results of this research generally support the previous findings that there is a weak but positive association between noun complement structure and L2 writing proficiency (Lan, Lucas, & Sun, 2019; Staples & Reppen, 2016; Taguchi et al., 2013) and that less advanced students use a restricted range

of controlling nouns (Jiang, 2015; Parkinson, 2013; Staples & Reppen, 2016). In addition, the change in the choice of head nouns by the student writers as proficiency advanced shows that they were in the process of acquiring genre conventions. However, there was still a high reliance on pragmatically vague nouns, indicating the need to enhance their awareness of the argumentative genre and expand the associated lexical repertoire.

4.5 The Grammatical Complexity Features with Little Contribution to the Association

As discussed in Section 4.3, the residual analysis showed that premodifying adjectives and nouns had little to do with the association between L2 writing proficiency and grammatical complexity, counter to the expectation that phrase-level complexity features would be strongly associated with high writing proficiency. Likewise, the finding that finite complement clauses headed by verbs or adjectives were not negatively correlated with proficiency levels did not conform to the expectation. These deviations from the expected results were perhaps largely attributable to the shortcoming of the statistical analysis based purely on frequency counts, that is, no concern for accuracy or lexical performance. In other words, committing errors or recycling identical words in the realizations of a particular grammatical feature was not filtered from frequency counts. However, accuracy is one crucial dimension of grammar that can be measured and

one distinct area of L2 performance together with fluency and complexity (Bulté & Housen, 2012; Norris & Ortega, 2009; Rimmer, 2006). Furthermore, the need to measure lexical performance in addition to CAF constructs should be taken into account for a more global and robust estimation of L2 performance (Skehan, 2009). The role of lexical diversity as a proficiency-level indicator was well empirically proven when comparing L2 writers with differing proficiency levels (e.g., Crossley et al. 2012; Grant & Ginther, 2000; Ishikawa, 2015) or comparing L1 and L2 English writers (e.g., Friginal et al., 2014; Staples & Reppen, 2016). While recognizing the interrelationship between grammar and vocabulary in writing development, Biber et al. (2011) also utilized lexico-grammatical information from Biber et al. (1999) in postulating the developmental sequence. Thus, this section takes up a close investigation into the actual use of the features found to be loosely associated with L2 writing proficiency, discussing how accuracy and lexical performance enable a more plausible explanation concerning the variations in the use of complexity features between proficiency levels.

4.5.1 Premodifying Adjectives

In contrast to prepositional phrases as nominal postmodifiers, specifically *of* phrases, premodifying adjectives displayed less clear-cut differences in frequencies between writing proficiency levels, despite being phrasal modifiers as NP constituents. Consistent with the present study, however, Biber et al. (2011)

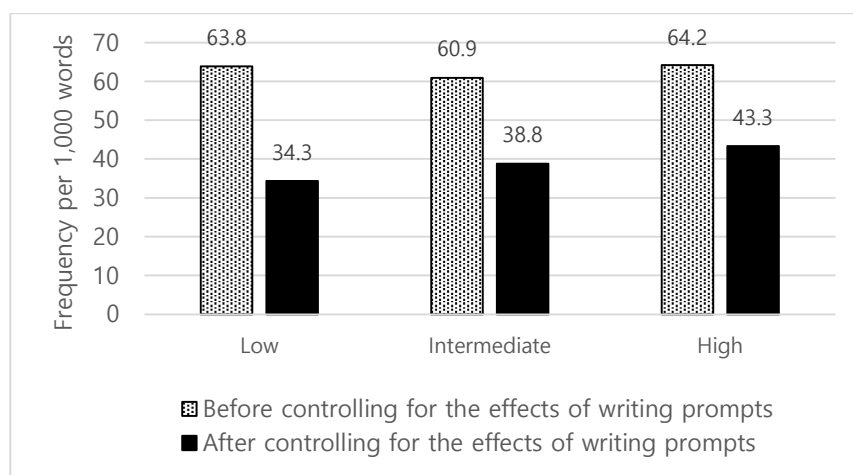
noted that as opposed to postmodifying prepositional phrases, which are much more frequent in written discourse than in conversational discourse, attributive adjectives made no noticeable difference since they are a single word. Their findings possibly suggest that premodifying adjectives do not reliably reflect proficiency-level differences as much as postmodifying prepositional phrases as they did in the present study.

The values of the adjusted standardized residuals indicate a weak negative correlation between L2 writing proficiency and the use of premodifying adjectives. To be specific, the small positive adjusted residuals associated with the students with low proficiency (1.6) and intermediate proficiency (0.6) indicate that there were moderately more premodifying adjectives in their essays than would be expected, whereas the small negative adjusted residuals associated with the students with high proficiency (-1.6) indicate that there were a little less than would be expected.

The qualitative check followed by the unexpected results of the residual analysis revealed that lexical diversity within this grammatical structure varied considerably depending on L2 writing proficiency. In fact, the highly advanced students used a wider range of premodifying adjectives, while the less advanced students drew heavily on adjective-noun sequences presented in writing prompts. To probe the influence of writing prompts, the frequencies of the six frequently occurring adjective-noun sequences, estimated to be related to the essay topics (i.e.,

cellular phone, military service, real name, public building, medical experiment, physical punishment) were calculated. Results indicated that the large majority of adjective-noun sequences in student writing was influenced by the essay prompts: 46.3% of the total adjective-noun sequences in the low-rated essays, 36.4% in the mid-rated essays, and 32.5% in the high-rated essays turned out to fall under the writing prompts. Thus, strikingly different results were yielded in the distributions of premodifying adjectives across proficiency levels when holding constant the effects of writing prompts, as depicted in Figure 4.4. In other words, after excluding the six adjective-noun sequences presented in writing prompts, the frequencies of premodifying adjectives showed a gradual increase as proficiency developed, unlike before controlling for the effects of writing prompts.

FIGURE 4.4 Frequencies of Premodifying Adjectives



Likewise, when adjusted residuals were re-calculated based on the frequencies minus topic-influenced premodifying adjectives, quite a different pattern emerged with respect to the value types: the mid-proficiency group (0.3) and high-proficiency group (0.2) had a positive adjusted value while the low-proficiency group (-0.7) had a negative adjusted value. The results after removing the topic effects also showed no significant association between premodifying adjectives and L2 writing proficiency but were parallel to Taguchi et al. (2013) and Lan, Lucas, and Sun (2019) in that a positive association was demonstrated.

The following excerpts illustrate these findings. Excerpt 3 (a) from a low-rated essay contains thirteen attributive adjectives in the 210-word essay. Notably, *cellular phone*, which seemed to be influenced by the writing prompt, appeared six times, accounting for nearly half the attributive adjectives used in this excerpt. Besides, the adjective *big* was used three times to modify different head nouns such as *accident* and *deal*.

Excerpt 3 (a) <Low-rated essay, file no. 748>

I think every drivers have to not allow their *cellular* phones. Car drivers, bike drivers, bicycle drivers have to. If you have lake of attention, you can have accidents. While driving uses *cellular* phone, very dangerous, and you can kill another people. Car accidents usually very *big* accident, so you have a car accident you maybe can` t walk or can` t go to bath yourself. It is very *big* deal. So when you driving you have to focus to driving one thing and should not use *cellular* phone. It is very *important* thing. It is same at bicycle drivers. People usually think bicycle is *safe* ride. But it is wrong. You have to same care with bicycle. Bicycle accidents are also very *big* and *dangerous* accidents. If you want to use your *cellular* phone,

stop driving and park your car at parking space. Then, you can use your phone. It is **best** way to use your phone at car. Car users maybe complain about that but for car users and for walkers it is very important. If drivers of automobiles use their **cellular** phones while driving they will take many people`s -walker`s-blames. You must have a responsibility. Driver`s **cellular** phone use must set of law.

By contrast, in Excerpt 3 (b) from a high-rated essay, which contains nineteen attributive adjectives in the 304-word essay, *cellular phone* appeared only twice, and the writer used alternative expressions in its steads such as *cell phone* and *mobile phone*. In the same vein, the writer showed a preference to use various near synonyms rather than repeating identical words, as evidenced by the use of *primary*, *main*, and *key* or the use of *deadly* and *serious*.

Excerpt 3 (b) <High-rated essay, file no. 3090>

Since talking to **cellular** phones distracts people to concentrate, using phones while driving should be banned. First of all, cell phone use increases the incidence of **deadly** car accidents. Many people get injured or dead due to car accidents everyday. Prevention of these accidents should be a **primary** goal of our society. Using **mobile** phones while driving is the **main** cause of accidents because it makes hard for drivers to focus on traffic signs. For example, a driver who are using the cell phone can ignore a **red** light and hit pedestrians on crosswalks. One mistake of a driver can put not just the man but **other** people in **serious** danger. Regulation of cell phone use while driving would help to reduce drivers' mistake and car accidents. If this action can save at least one life, it's worth it. Besides, cell phone use in driving can provoke **environmental** damages too. Traffic congestion is suggested to be a **main** source of air pollution. Automobiles emit **harmful** gases, and the longer people run their cars the more emissions are made. Talking to **cellular** phones slows down drivers' reaction. Therefore the congestions get worse and the **running** time of automobiles also increases. Although it seems too small to say cell phone use is a **key** factor of air pollution, the effect of reducing

traffic congestion is huge. There are several cases that succeeded to cut *enormous* amount of CO2 emission by enhancing traffic systems. If drivers don't use their *mobile* phones, there would be less congestions and it would help save the earth. Driving automobiles is convenient. However, if the drivers don't concentrate, it can bring about *serious* accidents and *environmental* destruction. As a conclusion I strongly think that *mobile* phone use while driving should be regulated, so that these problems can be prevented.

These observations confirm the finding of previous studies that lexical diversity is a distinctive feature of highly rated essays (Friginal et al., 2014; Reppen, 2001). In this regard, greater use does not necessarily represent greater development and what less advanced L2 writers specifically need is to produce diversity within NPs (Staples & Reppen, 2016). It means that writing development cannot be assessed solely by more or less usage of particular grammatical features and a more accurate picture of L2 writing development might be presented by adopting a comprehensive approach considering both grammatical and lexical aspects. In this view, the highly rated essays in the present study showed greater proficiency in using premodifying adjectives given their diverse lexical realizations of these structures.

Another noticeable contrast among the three proficiency groups is the use of NPs with multiple premodifying adjectives. AAN sequences were relatively rare in the low-rated essays, but these structures became more frequent as proficiency increased, and even AAAN sequences occurred only in the high-rated essays though they were unusual.

AAN sequences in the low-rated essays: ²⁵

cute brown rabbit, scary physical punishment, nonrational physical punishment, real personal information, short automagical answer

AAN sequences in the mid-rated essays:

serious physical punishment, moderate physical punishment, hopeful global world, best institutional treat, national military level, special political situation, strong military force, one-way physical punishment, mandatory military service, other week guys, current military system, suitable physical punishment, young Korean men, compulsive military service, lower national power, other mental punishment, only mental punishment, well-known scientific knowledge, other non-smoking people, current military system, healthy Korean men

AAN sequences in the high-rated essays:

basic human rights, unalienable human rights, clean public buildings, excessive physical punishment, moderate physical punishment, so-called bad students, higher medical technology, low medical technology, other better ways, other developed country, strong military power, strong military forces, strong national power, natural national duty, numerous positive outcomes, mandatory military duty, mandatory military service, compulsory military service, compulsory military duty, other serious problem, single public building, democratic free world, numerous dead bodies, holy Roman emperor, possible drastic consequences, unmeasurable economic service, other innocent people, young soon-to-be members, various social problems, only divided nation, recent military attacks, similar nonphysical systems, selective military service, primitive teaching method, small financial company, modern Korean education, other creative ways, tiny physical encounters, other chemical additives

AAAN sequences in the high-rated essays:

current mandatory military service, numerous other historical people

²⁵ Given the first-year college students' status as developing writers, those sequences judged comprehensible by a native English speaker were included in the list, although they contained mistakes as to lexical choices or semantic redundancy.

In summary, the qualitative analysis showed that considering multiple factors above frequency provides a more satisfactory account of differences in the use of premodifying adjectives between proficiency levels. In particular, the large proportion of the premodifying adjectives in the low-rated essays was influenced by writing prompts, and the low-proficiency students showed a marked tendency to recycle identical adjective-noun sequences. By contrast, the higher-rated essays showed greater lexical diversity in favor of using synonyms rather than repeating the exact words. Lastly, the use of a single adjective as prenominal modifier was increasingly extended to permit two or even three premodifying adjectives as proficiency improved.

4.5.2 Nouns as Nominal Premodifiers

As with premodifying adjectives, nouns as nominal premodifiers offered no substantive contribution to the correlation between grammatical complexity and writing proficiency levels. However, the value types of adjusted standardized residuals of premodifying nouns followed the exact opposite pattern from those of premodifying adjectives across proficiency levels: premodifying nouns were used a little less than expected in the low-rated essays (-0.3) and the mid-rated essays (-1.3), while a bit more than expected in the high-rated essays (1.5).

Although the magnitude of the residuals showed no obvious relation between the use of premodifying nouns and the three proficiency levels, the qualitative

review of student writing was revealing. A first finding was that there were no topic effects on noun-noun sequences, which could be one possible explanation for the opposite outcomes of the value types associated with premodifying adjectives and nouns in the low-rated essays. As discussed in the preceding section, a significant proportion of the adjective-noun sequences in the low-rated essays was produced under the influence of the writing prompts, and their repeated use resulted in the highest positive adjusted residuals of the low-proficiency group among the three groups, but this was not the case for noun-noun sequences. In fact, there was only one noun-noun sequence (i.e., *car accidents*) that repeatedly occurred between the low-rated essays, and even this was used in only two essays. The high-rated essays, on the other hand, showed several noun-noun sequences in which different head nouns are modified by a single topic-related noun. For example, Excerpt 4 (a) from a high-rated essay on the topic of using animals in medical experiments showed the repeated use of the premodifying noun *animal* in phrases such as *animal rights* and *animal experiments*; premodifying nouns and their head nouns are in bold italics.

Excerpt 4 (a) <High-rated essay, file no. 2516>

Those assertions about far too much ***animal rights*** are totally based on emotional response, and they simply treat animals as 'living', overlooking plants or other types of creatures because they don't show any action. Therefore, there is actually no need to be particularly sensitive to animals used in pharmaceutical labs. What we truly need is to find an appropriate point between human advantage and ***animal rights***, not totally banning

animal experiments. It is so manifest that *animal experiments* can't be abandoned considering their importance.

Likewise, in Excerpt 4 (b) from a high-rated essay on the topic of using cellular phones while driving, the premodifying noun *traffic* was found in phrases such as *traffic sign*, *traffic congestion*, and *traffic system*. In addition, the phrases such as *cell phone (use)*²⁶ and *mobile phone use* were used five times throughout the excerpt as a substitution for *cellular phone* to avoid its reuse. Excerpt 4 (c) from a mid-rated essay also showed a great number of *cell phone* and *phone call* used in place of *cellular phone*, which is the expression presented in the writing prompt.

Excerpt 4 (b) <High-rated essay, file no. 3090>

Using mobile phones while driving is the main cause of accidents because it makes hard for drivers to focus on *traffic signs*. For example, a driver who are using the *cell phone* can ignore a red light and hit pedestrians on crosswalks. One mistake of a driver can put not just the man but other people in serious danger. Regulation of *cell phone use* while driving would help to reduce drivers' mistake and *car accidents*. If this action can save at least one life, it's worth it. Besides, *cell phone use* in driving can provoke environmental damages too. *Traffic congestion* is suggested to be a main source of *air pollution*. Automobiles emit harmful gases, and the longer people run their cars the more emissions are made. Talking to cellular phones slows down drivers' reaction. Therefore the congestions get worse and the running time of automobiles also increases. Although it seems too small to say *cell phone use* is a key factor of *air pollution*, the effect of reducing *traffic congestion* is huge. There are several cases that

²⁶ In terms of noun-noun sequences, the distinction between compound nouns and syntactic construction could not be accurately determined in all instances. Thus, the sequences automatically extracted as noun-noun construction were all included in the analysis to exclude subjective judgment.

succeeded to cut enormous amount of CO2 emission by enhancing *traffic systems*. If drivers don't use their mobile phones, there would be less congestions and it would help save the earth. Driving automobiles is convenient. However, if the drivers don't concentrate, it can bring about serious accidents and environmental destruction. As a conclusion I strongly think that *mobile phone use* while driving should be regulated, so that these problems can be prevented.

Excerpt 4 (c) <Mid-rated essay, file no. 3276>

Using *cell phone* while driving is very dangerous, so drivers should not be allowed to use a *cell phone*. When driver uses a *cell phone*, he can't check *side mirrors* or a *room mirror*. If you don't check another cars from mirrors, it can take a *car accident*. It is why using *cell phone* while driving is dangerous. If driver really needs a *phone call*, he should use a headset. But using a headset is also dangerous action. So whenever driver needs a *phone call*, don't using a *cell phone* while driving is better than getting a *car accident*.

The four phrases, namely, *animal rights*, *animal experiments*, *cell phone*, and *phone call*, accounted for 19.0% and 20.5% of the total noun-noun sequences in the mid- and high-rated essays, respectively, whereas none of these phrases were found in the low-rated essays, reflecting a limited vocabulary range of the low-proficiency students for the productive use of this grammatical structure. In other words, the differences in the capability to leverage the topic-related vocabulary could be a reasonable explanation behind the positive adjusted residuals associated with the more proficient students but the negative adjusted residuals associated with the less proficient students.

The finding that the errors in the use of premodifying nouns decreased towards the higher proficiency levels also indicates that the high-proficiency group was

more familiar with these phrasal structures. There were three prominent types of errors, which the less proficient students mostly made. The first type was to use plural nouns in place of singular or possessive nouns as in Examples (36) and (37). The second type was to use premodifying nouns in cases where a prepositional phrase should be used as in Examples (38) through (40). In each of these examples, the correct expressions are *a factor in the crash*, *the advance in medical science*, and *the spirit of soldiers*. These two types of errors occurred relatively frequently in the low- and mid-rated essays, whereas none of these errors were found in the high-rated essays. The third type of error was to make the wrong choice of lexical categories, particularly using nouns in place of adjectives, as in Examples (41) through (43). This last type of error was also found in the high-rated essays, although not as frequent.

- (36) So, this experiment get the problem of *animals life*. <Low, file 66>
- (37) So experiments that are important and crucial to *humans health* can retail animal's sacrifice. <Mid, file 1062>
- (38) I think that this *crash fator* consist of three. <Low, file 661>
- (39) But If animals can't be used in experiments, humanbeings medical *science advance* will be slowed down. <Mid, file 1062>
- (40) *Soldier spirit* is just for soldiers, not for citizens. <Mid, file 1386>
- (41) If smokers want to smoke, they go to the *privacy rooms*. <Low, file 3172>
- (42) By not using cell phones in car, we can save money and enjoy prolonged *health lives*. <Mid, file 2540>
- (43) It is really important that people share the opinions and debate each other in *democracy society*. <High, file 2178>

These three types of errors occurred eight times in the low-rated essays

compared with correct use 43 times (15.7%), fifteen times in the mid-rated essays compared with correct use 126 times (10.6%), and only three times in the high-rated essays compared with correct use 196 times (1.5%). Considering that “when writers try for greater complexity than they are perhaps comfortable with, errors are more likely to occur” (Parkinson & Musgrave, 2014, p. 56), we can infer that the high-proficiency students reached the developmental stage of making comfortable use of premodifying nouns while the lower-proficiency students have not yet passed this stage.

In sum, despite little correlation between the occurrence rate of premodifying nouns and L2 writing proficiency, it was observed from the qualitative analysis that the high-proficiency students were more proficient in their use, given the high lexical diversity associated with these structures as well as very few errors in their formation. By contrast, the writing samples of the lower-proficiency students illustrated the need to widen their vocabulary range for the diverse lexical realizations of noun-noun structures and improve accuracy by addressing three types of recurrent errors.

4.5.3 Finite Complement Clauses Controlled by Verbs or Adjectives

The smallest sum of absolute values of adjusted residuals related to finite complement clauses after verbs or adjectives indicates that these structures

contributed the least to the association. Specifically, the small positive adjusted residual values associated with the low-proficiency students (1.1) and high-proficiency students (0.6) indicate that they used moderately more *that*-clauses controlled by verbs or adjectives than would be expected, whereas the small negative adjusted residual value associated with the intermediate-proficiency students (-1.3) indicates that they used a little less than would be expected.

As in the case of premodifying adjectives and nouns, the residuals for the cells of *that*-clauses headed by verbs or adjectives did not differ significantly among the groups. Further examination of controlling verbs, however, revealed a significant contrast in the degree of lexical diversity. Table 4.6 draws a comparison of the normed frequencies of the top six most common controlling verbs in the three sets of writing.

TABLE 4.6
Top 6 Verbs Controlling *That*-clauses

	Low	%	Intermediate	%	High	%
<i>Think</i>	426.3	62.8	209.7	39.8	104.8	16.9
<i>(Dis)agree</i>	110.5	16.3	40.9	7.8	16.8	2.7
<i>Say</i>	15.8	2.3	92.1	17.5	104.8	16.9
<i>Believe</i>	0	0.0	30.7	5.8	75.4	12.2
<i>Insist</i>	15.8	2.3	40.9	7.8	54.5	8.8
<i>Argue</i>	0	0.0	10.2	1.9	54.5	8.8

Note. Normed per 100,000 words.

The notably frequent verb in the low-rated essays was *think* (62.8%) and the

second most frequent verb was *(dis)agree* (16.3%). These two extremely frequent verbs made up 79.1% of the total controlling verbs used by the low-proficiency students with the other nine controlling verbs (e.g., *say*, *insist*, *know*) occurring only once, reflecting their very limited set of verbs taking complement clauses. The intermediate-proficiency students used a wider variety of verbs to create these structures. *Think* (39.8%) was the most common verb in the mid-rated essays as well but to a lesser extent than in the low-rated essays. The second most frequent verb was *say* (17.5%), followed by *(dis)agree* (7.8%), *insist* (7.8%), and *believe* (5.8%). Each of the remaining twelve controlling verbs in the mid-rated essays (e.g., *suggest*, *know*, *mean*) had an occurrence rate lower than 3% of the whole controlling verbs. The high-rated essays showed the greatest lexical diversity in verb complement clauses. The verbs *think* and *say* (16.9% each) were used most frequently, and the other relatively common verbs included *believe* (12.2%), *insist* (8.8%), and *argue* (8.8%). Each of the remaining twenty-four controlling verbs found in the high-rated essays (e.g., *suggest*, *show*, *realize*) had a frequency lower than 4%.

In addition to the increasingly balanced use of various controlling verbs as proficiency developed, Table 4.6 also shows a non-academic way of using verb complement clauses even at high proficiency. The highly favored controlling verbs in student writing such as *think*, *say*, and *believe* are extremely widespread controlling verbs in conversation: *think* is by far the first, *say* is the second, and

believe is the seventh most used controlling verb in conversation according to Biber et al.'s (1999) corpus-based findings. Among these three, the verbs *think* and *believe* representing “a mental process of cognition” (Halliday & Matthiessen, 2004, p. 451) are likely to impair information with subjectivity and uncertainty by emphasizing one’s beliefs and opinions (Parkinson, 2013). The extensive use of mental verbs in student writing thus seems entirely contrary to academic writers’ general intention of using controlling verbs to “raise the authority of the contents of the *that*-clause” (Parkinson, 2013, p. 429). Examples (44) through (48) illustrate how students drew heavily on personal thoughts and beliefs to support their argument, although a writer’s argument should be supported by relevant facts or data, not private opinions or feelings (Qin & Karabacak, 2010). It is for this reason that mental verbs are far less favored by academic writers than communication verbs (e.g., *suggest*, *indicate*, *show*) (Biber et al., 1999).

- (44) I **think** this is very difficult problem. <Low, file 1099>
- (45) I **think** this act is not fair. <Mid, file 2389>
- (46) First, I don’t **think** that will work. <High, file 2624>
- (47) I don’t **believe** that anybody feels the way I do about you. <Mid, file 2253>
- (48) Though there can be many opinions about this, I firmly **believe** that not all Koreans have to complete military service. <High, file 358>

Previous L2 writing research has reported similar distributional patterns of controlling verbs, observing the predominant use of verbs suited for conversation such as *think*, *believe*, and *say*, i.e., lexical choices that deviate from academic

norms (Back, 2011; Biber & Reppen, 1998; Huh & Hwang, 2011; Oh, 2007; Parkinson, 2013; Staples & Reppen, 2016). These findings align with the fact that *that*-clauses after highly prevalent verbs are placed in Stage 1 of Biber et al.'s (2011) framework assuming that “these relatively fixed lexico-grammatical combinations” are acquired early (p. 31), while a broader spectrum of verbs that control *that*-clauses are assigned to Stage 2 (see Table 2.2).

However, there seems to be another factor that induced the extensive use of mental verbs in student writing. In an argumentative task that requires students to take one position on the issue at hand and argue for their position, they are allowed to freely present their viewpoint and attitude to some degree. Toward this end, mental verbs (e.g., *think*, *believe*, *know*) can serve a stance-marking function specifically as “markers of beliefs” whereby a writer “acknowledges that assertions are the result of self or others’ beliefs” (Reilly et al., 2002; Uccelli et al., 2013, p. 15) or “express(es) the writer’s opinion of the validity of the following information” (Hyland & Tse, 2005a, p. 43). The high frequency verb *(dis)agree*, especially in the low-rated essays, may also be interpreted in light of the influence of genre characteristics. As students had to choose a particular side in argumentative writing, the phrase *I (dis)agree* seems to be frequently used to clarify their position as in Examples (49) through (51).

(49) I **agree** that people must use their real name on the Internet. <Low, file 3015>

- (50) But I **disagree** that all Korean men be forced to complete military service.
<Mid, file 1451>
- (51) Considering the side effect and ineffectiveness that outweigh benefits, I strongly **disagree** that people should post their ideas with real name.
<High, file 2564>

The controlling verbs relatively common in the high-rated essays but rare or non-existent in the mid- and low-rated essays were *insist* and *argue*. This points to little awareness of argumentative genre at lower levels given the importance of using “particular verbs more specific to the task of argumentation in writing” (Staples & Reppen, 2016, p. 18). In most cases, *insist* and *argue* controlling *that*-clauses co-occurred with third person plural subjects (e.g., *they*, *some people*) as in Examples (52) and (53) to state opposing arguments and indicate writers’ awareness of them, rather than developing their own views on the issue. This balanced approach might serve to make readers with alternative opinions engage in discussions and interact with the writer. Such a balanced and interactive style of writing identified in the high-rated essays is comparable to the findings of previous studies that advanced writers tend to communicate their ideas in a less confrontational manner by giving the audience more leeway for disagreement or alternative viewpoints, resulting in greater potential for dialogue with readers (Oh & Kang, 2013; White & Sano, 2006)

- (52) Some smokers **argue** that the law that prohibit to smoke in public places is unnecessary and that its’ their rights to decide where they smoke. <High, file 2513>

(53) They **insist** that spending almost two years in army is waste of time.
<High, file 32>

With respect to communication verbs, only three kinds were found in each of the high- and mid-rated essays: *suggest*, *show*, and *prove* in the high-rated essays, and *suggest*, *show*, and *reveal* in the mid-rated essays, while no communication verbs were attested to in the low-rated essays. The three verbs made up only around 9% and 6% of the total controlling verbs used in the high-and mid-rated essays, respectively. The occurrence rate of communication verbs taking *that*-clauses in student writing seems largely different from that in typical academic writing where these types of verbs are commonly found with “a non-personal subject to report stance that is not overtly associated with the thought or feelings of human observers” (Biber et al., 1999, p. 670). The discrepancy between the findings of the present study and Biber et al. (1999) can largely be attributable to the difference in the specific academic genre examined. To be specific, argumentative essays investigated in this study may differ in terms of required lexical items from research articles analyzed by Biber et al. (1999) where “evidential verbs” (e.g., *suggest*, *speculate*) are used in great abundance to “specify and acknowledge previous findings, and also take a stance toward those findings by referring to either speculative or deductive judgments” (Hyland, 1996, pp. 266-267).

A contrast between mid- and high-proficiency groups in the use of

communication verbs was that the high-proficiency students favored abstract subjects such as *results* or *examples* as in Examples (54) through (56) rather than human subjects such as *I* or *people* as in Examples (57) and (58), but the opposite was true for the intermediate-proficiency students. This reflects the acquisition of some upper-level students of how academic language uses communication verbs taking *that*-clauses, given that academic writers often use abstract subjects in these structures to maintain objectivity by not attributing the source of the *that*-clause information to themselves or others (Biber et al., 1999; Charles, 2006; Hyland & Tse, 2005b).

- (54) This attack **suggests** that North Korea can take more actions and we can get serious damage from them. <High, file 32>
- (55) The result **showed** that the group with the physical punishment had a less level of students misbehaving than the other without the physical punishment. <High, file 2318>
- (56) This example **proves** that animals for medical experiments are helpful for mankind. <High, file 1619>
- (57) The people who don't agree with the physical punishment **suggest** that the physical punishment will only hurt children't heart and that words can solve the problem and make children think properly. <Mid, file 393>
- (58) Therefore, I **suggest** that the military should be adopt apply system which is adopted in the U.S. <Mid, file 2544>

In summary, although the frequencies of verb complement clauses across the three groups bore little relevance to L2 writing proficiency, the comparison of controlling verbs used in the three subcorpora suggested that verb complement clauses in the higher-rated essays were associated with a broader range of

controlling verbs than those in the lower-rated essays. Besides, some high-rated essays demonstrated an interactive style of writing, using these structures to address conflicting claims on the subject and provide a more balanced perspective. Yet, qualitative reviews also showed that even highly proficient students relied heavily on controlling verbs characteristic of conversational language rather than drawing on high frequency academic verbs, though effects of genre (i.e., argumentative essays) seemed to be involved. Only a few high-rated essays demonstrated an academic way of using verb complement clauses along with abstract subjects. These findings suggest that students at all proficiency levels need further development on the use of verb complement clauses by adopting more academic controlling verbs and non-human subjects for less overt attribution of stance.

Complement clauses headed by an adjective were by far less common than those headed by a verb in student writing, which aligns with prior corpus-based studies (Biber et al., 1999; Biber et al., 2011; Parkinson, 2013). Moreover, they had the lowest frequencies at all proficiency levels among the seventeen features analyzed. There were only five instances of *that*-clauses controlled by adjectives in the entire sets of writing: three in the high-rated essays, two in the mid-rated essays, and none in the low-rated essays. In these five occurrences, three adjectives were used to control *that*-clauses: *sure*, *convinced*, and *afraid*. These certainty and affective adjectives typically co-occur with a human subject, overtly attributing

the feeling to that person as shown in Examples (59) through (61) (Biber et al., 1999, p. 672).

- (59) And I'm **sure** that they are right. <Mid, file 2389>
- (60) However, I'm **convinced** that physical punishment should not be allowed in all schools. <High, file 493>
- (61) They said they were **afraid** that the teacher would know who was do that and would be angry again. <High, file 2178>

Such explicit attribution of personal feelings in post-predicate complement clauses makes these constructions particularly common in conversation rather than academic writing. Based on the register distribution, post-predicate *that*-clauses were hypothesized to be acquired early, whereas extraposed *that*-clauses, which occur in academic writing with relatively high frequency due to their implicit attribution of stance, were argued to be acquired developmentally late (Biber et al., 1999; Biber, Conrad, & Leech, 2002; Biber et al., 2011). In this regard, an unexpected finding of this study was that extraposed *that*-clauses controlled by adjectives occurred with higher frequency in student writing than post-predicate ones, though extraposed constructions were not frequent in absolute terms. There were sixteen instances of extraposed *that*-clauses in the entire sets of writing: twelve in the high-rated essays, four in the mid-rated essays, and none in the low-rated essays. Of these sixteen occurrences, the adjective *true* was used six times, *unfair* and *important* twice, and a few other adjectives (i.e., *clear*, *likely*, *sure*, *certain*, *obvious*, *real*) only once. See Examples (62) through (64).

- (62) Obviously, it is **true** that excessive physical punishment should not be allowed. <High, file 2132>
- (63) But it actually is **unfair** that only men should lose two years. <High, file 2044>
- (64) It is really **important** that people share the opinions and debate each other in democracy society. <High, file 2178>

The relatively higher frequency of extraposed complement clauses compared to post-predicate ones in student writing, contrary to Biber et al.'s (2011) developmental stages, is presumably attributable to the fact that most Korean L2 learners are taught written English and relatively less familiar with spoken English. As with post-predicate *that*-clauses, however, extraposed ones following adjectives were of extreme rarity in absolute terms at all proficiency levels. Given that extraposed *that*-clauses, especially headed by “necessity or importance adjectives” such as *essential* and *vital*, are quite common in academic texts (Biber et al., 1999, p. 674), students need to be familiarized with these lexicogrammatical patterns through focused exposure.

CHAPTER 5. CONCLUSION

This chapter firstly summarizes the key findings of this research in Section 5.1 by providing answers to the research questions. Then, based on these findings, pedagogical implications for L2 writing instruction are discussed in Section 5.2. Lastly, limitations of this research that provide suggestions for future work in this area are presented in Section 5.3.

5.1 Major Findings

This study sought to estimate the developmental trajectory of grammatical complexity for L1 Korean writers by investigating the association between the use of complexity features and L2 writing proficiency. For this purpose, complexity features used in argumentative essays of Korean college students with differing proficiency levels (i.e., low, intermediate, high) were examined. For measures of grammatical complexity, this study applied the developmental framework put forth by Biber et al. (2011), which is based on the shift from clausal to phrasal complexity. The nine clausal and eight phrasal complexity features from Biber et al.'s (2011) index were used to operate grammatical complexity in student writing. In what follows, the two major findings via a Pearson Chi-square test and residual analysis are summarized.

Firstly, the finding regarding the relationship between L2 writing proficiency

and grammatical complexity confirms previous studies (Biber et al., 2011; Kim, 2020; Kyle & Crossley, 2018; Lan, Lucas, & Sun, 2019; Parkinson & Musgrave, 2014; Taguchi et al., 2013), showing that the three proficiency levels are associated with the frequencies of occurrence of clause- and phrase-level complexity features.

Second, and even more importantly, the finding regarding the major contributors to the association supports the developmental sequence hypothesized by Biber et al. (2011), to the extent that finite adverbial clauses (Stage 2) were the key features of the low-rated essays; WH relative clauses (Stage 3) were the most prominent complexity devices of the mid-rated essays; and *of* phrases as postmodifiers (Stages 3-5) best represented the complex styles of the high-rated essays. The observed developmental sequence of Korean college students shows a stepwise increase in grammatical complexity in terms of two parameters: structural form and syntactic function. To be specific, finite adverbial clauses associated with the low-rated essays represent the most basic complexity as “finite dependent clauses functioning as constituents in other clauses” (ibid., p.29); WH relative clauses associated with the mid-rated essays reflect the moderate complexity, where the two parameters are mixed, as “finite clause types functioning as a constituent in a noun phrase” (ibid., p. 27); and *of* phrases associated with the high-rated essays bear on the highest degree of complexity as “phrasal (nonclausal) dependent structures that function as constituents in noun phrases” (ibid., p. 30).

Despite the fact that noun complement clauses are a mixture of two parameters as with WH relatives, they did not fit the intermediate stage due to their primary functions of signaling stance in academic writing. In fact, none of the three groups showed the especially frequent use of noun complement clauses, and their use was restricted to a narrow range of controlling nouns. These findings suggest that noun complement clauses are the most advanced complexity features, as shown in their corresponding developmental stage in Biber et al.'s (2011) model.

An unexpected finding was that premodifying adjectives and nouns were not significantly associated with L2 writing proficiency despite being noun-modifying phrasal features typical of advanced writing. The qualitative analysis of student writing, however, provided the three main sources of variation across different proficiency levels. First, a large proportion of adjective-noun sequences in the lower-rated essays were influenced by the writing prompts. Second, the less advanced students showed the repetitive use of particular phrases due to their limited range of vocabulary, whereas the more advanced students preferred the use of near synonyms rather than recycling the identical sequences, showing greater lexical diversity associated with grammatical structures. The lexico-grammatical resources of the highly proficient students even enabled the greater production of NPs with multiple premodifiers. Third, the number of errors in the use of nominal premodifiers showed an evident trend of decreasing as proficiency improved. Taken together, the findings of the qualitative analysis suggested that phrasal noun

modifiers were complexity features that the higher-proficiency students have largely acquired, while the lower-proficiency students were still not comfortable with.

The finding that the frequencies of verb complement clauses located at the early developmental stages of Biber et al.'s (2011) model made no marked difference among proficiency levels also deviated from the initial expectation that students with lower proficiency would show the predominant use of clausal features typical of conversation. The qualitative analysis, however, revealed that the more proficient students showed moderately increased lexical diversity in the choice of controlling verbs, but somewhat surprisingly, even the highly advanced students relied on verbs characteristic of conversational and non-academic language.

The overall trend was that the grammatical complexity of student essays was increasingly aligned with academic writing norms, considering the greater use of phrasal features functioning as NP constituents as proficiency advanced. The qualitative analysis demonstrated, however, that the development of grammatical complexity is inextricably connected and should be considered together with lexical realizations and accuracy.

5.2 Pedagogical Implications

This research demonstrated a significant association of grammatical

complexity with L2 writing proficiency, largely validating Biber et al.'s (2011) hypothesized developmental progression. More importantly, it turned out that the association is derived primarily from the four complexity features (i.e., finite adverbial clauses, WH relative clauses, postmodifying *of* phrases, finite complement clauses controlled by nouns) to which particular attention should be paid. Additionally, the qualitative analysis revealed critical areas of language instruction in terms of accuracy and lexico-grammatical choices.

The overall results provide clear instructions on how to assess writing performance and customize writing courses accordingly to facilitate improvement in English composition skills in the context of formal academic writing. First, the empirically derived developmental stages for complexity features in the present study may help develop more detailed rating scale descriptors for grammatical complexity of L2 writers' performance by explicitly defining prominent grammatical structures displayed by L2 writers at a certain level of development (Knoch, 2009). The more detailed descriptors may then generate detailed feedback on the use of complexity features, which can, in turn, encourage students to use a further advanced level of grammatical structures, provided that they are developmentally ready for that level of complexity. Second, classroom focus on the academically oriented lexical realizations of grammatical structures may be of value for advanced as well as less advanced students to speed up their acquisition of academic writing norms. This is particularly the case for the choice of

controlling nouns and verbs taking *that* complement clauses, given that the controlling words even in highly rated essays were very limited and largely reflected conversational norms. Third, language instruction should also deal with problems of accuracy with which grammatical structures are used. In this regard, particular attention can be paid to relative clauses and premodifying nouns where a few recurrent errors were found.

The need for focused instructional treatments on particular grammatical structures draws on the two findings of research synthesis and meta-analysis on the effectiveness of L2 instruction: first, explicit types of treatment showed larger effect sizes than implicit type treatments, and second, the effects were even durable (Norris & Ortega, 2001). In other words, L2 instruction can raise the effect of education by inducing learners to “give deliberate attention to language features” and to “process the language features in deep and thoughtful ways” (Nation, 2007, p. 6). One way of providing such opportunities is using consciousness-raising activities designed to “develop an awareness of the form, function, and meaning, and use of complex nominal groups at the level of explicit knowledge” (Musgrave & Parkinson, 2014, p. 154). To illustrate, students can be required to compare writings from low- and high-proficiency groups or writings from student writers and professionals and then identify the differences in grammatical structures at clausal and phrasal levels. Along the same lines, comparing the high frequency verbs and nouns controlling *that*-clauses in student writing and advanced academic

prose may allow students to notice the differences and equip themselves with a wide selection of academic controlling words. Such exposure to the use of complex grammatical constructions might help students inductively capture the characteristics typical and atypical of professional academic writing. The focused instruction is ultimately expected to help student writers become familiar with targeted grammatical structures and make greater use of more advanced phrasal modifiers in their own writing, by virtue of which their composition will take on more professional academic writing.

5.3 Limitations and Prospect for Future Research

This study has four main limitations. Firstly, this study only examined grammatical features that can be tagged by the CLAWS tagger. Thus, important structural variants involving the omission of the complementizer *that* in *that*-clauses headed by verbs or adjectives (e.g., *I think \emptyset physical punishment should be banned in all schools*) or the relativizers *that*, *who*, *which*, and *whom* in relative clauses (e.g., *physical punishment only makes children to be afraid of the pain \emptyset they get*) were not included in the analysis. The exclusion of structures with no complementizer in this study could be justified by the corpus finding that “retention of *that* is the norm in academic prose,” although several discourse and grammatical factors are involved in the choice of whether to omit or keep *that* in complement clauses (Biber et al., 1999, p. 680). Likewise, the zero relativizer is

proportionally much less frequent in academic texts despite the higher frequency of relative clauses in written discourse than in conversational discourse (Biber et al., 1999). Nevertheless, future studies will benefit from taking into consideration structures with no relativizer or complementizer in L2 writing to examine whether L2 writers are progressing toward the academic norm, given conflicting empirical findings on the relationship between the frequency of omissions and L2 writing proficiency (e.g., Ginther & Grant, 1997; Lan, Lucas, & Sun, 2019).

Second, as discussed in Section 2.4.3, written production with a specific time limit might not elicit heavy nominal groups typical of advanced academic prose (Ellis & Yuan, 2004; Lu, 2011). Thus, the language produced for the timed argumentative essays as in this study could fail to fully reflect individual grammatical proficiency while reducing the gap between developmentally more advanced and less advanced students in terms of the kinds and degree of complexity. In this regard, investigation on untimed academic writing may enable a more precise evaluation of complexity associated with different developmental levels.

Third, the corpus on which this study is based (i.e., YELC 2011) only contains writing samples from Yonsei University freshmen and may not be representative of the wider student body. In other words, the samples are heavily skewed and far more homogeneous than the target population since they were taken from a selected group of students with high academic ability at a major university. It thus

seems likely that the weak effect size obtained in this research might be, in part, due to “range restriction (also called truncated samples)” caused by “selecting only the top individuals in the score distribution” (Bachman, 2004, p. 96). Besides, as noted in the description of YELC 2011 in Section 3.1.1, students with the most advanced English were excluded under the exemption policy, which could further restrict the range of the study population and substantially attenuate the correlation between L2 writing proficiency and grammatical complexity (Sackett et al., 2002). The small size of the corpus, including only 78 essays per proficiency level, also made it difficult to partially compensate for the restricted range. Thus, future studies should employ a more representative corpus, which includes the full range of L2 writing proficiency, in order to be better able to generalize the research findings to the wider student population that the corpus originally intended to represent.

Fourth, due to practical concerns, this study compared the performance of L2 writers at different proficiency levels with the aim of elucidating their developmental patterns of grammatical complexities. Thus, the reported results of this study may not represent individual development over time. To provide a genuine picture of developmental trajectories, future studies might use longitudinal data collected from the same L2 writers over an extended period of time.

REFERENCES

- Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). Sampling: Why and how of it. *Indian Journal of Medical Specialties*, 4(2), 330-333.
- Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons.
- Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Ed.), *Automatic treatment and analysis of learner corpus data* (pp. 249–264). John Benjamins.
- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3), 91-93.
- Aktas, R. N., & Cortes, V. (2008). Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes*, 7(1), 3-14.
- Arnaud, P. J. L. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. In P. J. L. Arnaud, & H. Bejoint (Ed.), *Vocabulary and applied linguistics* (pp. 133–145). Macmillan.
- Arthur, B. (1979). Short-term changes in EFL composition skills. In C. A. Yorio, K. Perkins, & J. Schachter (Ed.), *On TESOL '79: The learner in focus* (pp. 330–342). TESOL.
- Aziz, R. A., & Don, Z. M. (2019). Tagging L2 writing: Learner errors and the performance of an automated part-of-speech tagger. *GEMA Online Journal of Language Studies*, 19(3), 140-155.
- Bachman, L. F. (2004). *Statistical analyses for language assessment book*. Cambridge University Press.
- Back, J. H. (2011). A corpus-based study of hedging in Korean EFL students'

- academic essays. *Studies in Modern Grammar*, 65, 279-307.
- Baik, M. J. (1994). Syntactic features of Englishization in Korean. *World Englishes*, 13(2), 155-166.
- Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, 26(2), 390-395.
- Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11(1), 17-34.
- Becker, A. (2010). Distinguishing linguistic and discourse features in ESL students' written performance. *Modern Journal of Applied Linguistics*, 2(6), 406-424.
- Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing*, 22(2), 185-200.
- Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Lawrence Erlbaum.
- Biber, D. (1985). Investigating macroscopic textual variation through multi-feature/multi-dimensional analyses. *Linguistics*, 23, 337-360.
- Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62, 384-414.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (1992). On the complexity of discourse complexity: A multidimensional analysis. *Discourse Processes*, 15, 133-163.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken*

- and written English*. Pearson Education Limited.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, 36(1), 9-48.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5-35.
- Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5), 639-668.
- Biber, D., Gray, B., Staples, S., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, 46, 100869.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). Longman.
- Biber, D., & Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. In S. Granger (Ed.), *Learner English on computer* (pp. 145–158). Longman.
- Brown, H. D. (2000). *Principles of language learning and teaching* (Vol. 4). Longman.
- Brown, A., Iwashita, N., & McNamara, T. (2005). An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tasks. *ETS Research Report Series*, 2005(1), i-157.
- Bruner, J. (1986). *Actual minds, possible worlds*. Harvard University Press.
- Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Ed.), *Dimensions of L2 performance*

- and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). John Benjamins.
- Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, 3(3), 179-201.
- Charles, M. (2006). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics*, 27(3), 492-518.
- Choe, J. W., & Song, J. Y. (2013). The topical classification of essays by college student English learners using hierarchical clustering. *Language Information*, 17(5), 93-115.
- Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research*, 69(5), 176-183.
- Crossley, S. A., & McNamara, D. S. (2012). Detecting the first language of second language writers using automated indices of cohesion, lexical sophistication, syntactic complexity and conceptual knowledge. In S. Jarvis, & S. A. Crossley (Ed.), *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 106-126). Multilingual Matters.
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66-79.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2), 243-263.
- Cumming, A., & Mellow, D. (1996). An investigation into the validity of written indicators of second language proficiency. In A. Cumming, & R. Berwick (Ed.), *Validation in language testing* (pp. 72-93). Multilingual Matters.
- Dvorak, T. (1987). Is written FL like oral FL? In B. VanPatten, T. Dvorak, & J. Lee (Ed.), *Foreign language learning: A research perspective* (pp. 79-91).

Newbury House.

- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26(1), 59-84.
- Fazilatfar, A. M., Fallah, N., Hamavandi, M., & Rostamian, M. (2014). The effect of unfocused written corrective feedback on syntactic and lexical complexity of L2 writing. *Procedia-Social and Behavioral Sciences*, 98, 482-488.
- Flahive, D. E., & Snow, B. G. (1980). Measures of syntactic complexity in evaluating ESL compositions. In J. W. Oller, & K. Perkins (Ed.), *Research in language testing* (pp. 171-176). Newbury House.
- Flowerdew, J. (2003). Signalling nouns in discourse. *English for Specific Purposes*, 22(4), 329-346.
- Flowerdew, J. (2006). Use of signalling nouns in a learner corpus. *International Journal of Corpus Linguistics*, 11(3), 345-362.
- Ford, C. E. (1997). Speaking conditionally: Some contexts for if-clauses in conversation. In A. Athanasiadou, & R. Dirven (Ed.), *On conditionals again* (pp. 387-413). John Benjamins.
- Friginal, E., Li, M., & Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing*, 23, 1-16.
- Gaies, S. J. (1976). Sentence-combining: A technique for assessing proficiency in a second language. Paper delivered at the Conference on Perspectives on Language. University of Louisville.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Ed.), *Corpus annotation: Linguistic information from computer text corpora* (pp. 102-121). Longman.
- Ginther, A., & Grant, L. (1997). The influence of proficiency, language

- background, and topic on the production of grammatical form and error on the Test of Written English. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Ed.), *Current developments and alternatives in language assessment* (pp. 385–397). University of Jyväskylä.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123-145.
- Gries, S. T., & Newman, J. (2013). Creating and using corpora. In R. J. Podesva, & D. Sharma (Ed.), *Research methods in linguistics* (pp. 257-287). Cambridge University Press.
- Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18(3), 218-238.
- Halliday, M. A. K. (1989). *Spoken and written language*. Oxford University Press.
- Halliday, M. A. K. (2004). *The language of science*. Continuum.
- Halliday, M. A. K., & Hasan, R. (1985). *Language, context and text: Aspects of language in social-semiotic perspective*. Deakin University Press.
- Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power* (Vol. first printed in 1993, reprinted in 1996). Falmer Press.
- Halliday, M. A. K., & Matthiessen, C. (1999). *Construing experience through meaning: A language-based approach to cognition*. Cassell.
- Halliday, M. A. K., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd ed.). Arnold.
- Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Newbury House Publishers.
- Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *The Modern Language Journal*, 80(3), 309-326.

- Hinkel, E. (2013). Research findings on teaching grammar for academic writing. *English Teaching*, 68(4), 3-21.
- Hirano, K. (1991). The effect of audience on the efficacy of objective measures of EFL proficiency in Japanese university students. *Annual Review of English Language Education in Japan*, 2, 21-30.
- Homburg, T. J. (1984). Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL Quarterly*, 18(1), 87-107.
- Ho-Peng, L. (1983). Using T-unit measures to assess writing proficiency of university ESL students. *RELC Journal*, 14(2), 35-43.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Hu, C., & Li, X. (2015). Epistemic modality in the argumentative essays of Chinese EFL learners. *English Language Teaching*, 8(6), 20-31.
- Huh, M. H., & Hwang, I. Y. (2011). Actual use of that-clauses in EFL writing. *The New Korean Journal of English Language & Literature*, 53(1), 267-290.
- Hunston, S., & Francis, G. (1999). *Pattern grammar*. Benjamins.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (Vol. 3). National Council of Teachers of English.
- Hyland, K. (1994). Hedging in academic writing and EAF textbooks. *English for Specific Purposes*, 13(3), 239-256.
- Hyland, K. (1996). Talking to the academy: Forms of hedging in science research articles. *Written Communication*, 13(2), 251-281.
- Hyland, K., & Tse, P. (2005a). Evaluative that constructions: Signalling stance in research abstracts. *Functions of Language*, 12(1), 39-63.
- Hyland, K., & Tse, P. (2005b). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes*, 24(2), 123-139.
- Ishikawa, T. (2007). The effects of increasing task complexity along the -/+ Here-and-Now dimension. In M. P. García Mayo (Ed.), *Investigating tasks in*

- formal language learning* (pp. 136–156). Multilingual Matters.
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4(1), 51-69.
- Ishikawa, S. I. (2015). Lexical development in L2 English learners' speeches and writings. *Procedia-Social and Behavioral Sciences*, 198, 202-210.
- Jang, J., & Rhee, S. C. (2014). A corpus-based study of the use of shell nouns as cohesive devices in Korean EFL writing. *English Language & Literature Teaching*, 20(3), 71-93.
- Jiang, F. K. (2015). Nominal stance construction in L1 and L2 students' writing. *Journal of English for Academic Purposes*, 20, 90-102.
- Jiang, J., Bi, P., & Liu, H. (2019). Syntactic complexity development in the writings of EFL learners: Insights from a dependency syntactically-annotated corpus. *Journal of Second Language Writing*, 46, 100666.
- Kameen, P. (1979). Syntactic skill and ESL writing quality. In C. Yorio, K. Perkins, & J. Schachter (Ed.), *On TESOL '79: The learner in focus* (pp. 343-364). TESOL.
- Kang, J. Y. (2005). Written narratives as an index of L2 competence in Korean EFL learners. *Journal of Second Language Writing*, 14(4), 259-279.
- Kang, S., & Oh, S. Y. (2022). A corpus-based study on the use of English relative clauses by Korean EFL learners of different proficiency. *The Journal of Asia TEFL*, 19(1), 1-16.
- Kern, R. G., & Schultz, J. M. (1992). The effects of composition instruction on intermediate level French students' writing performance: Some preliminary findings. *The Modern Language Journal*, 76(1), 1-13.
- Kim, C. H., & Suh, H. W. (2014). Epistemic rhetorical stance: Hedges and boosters in L1 and L2 students' English writing. *The Linguistic Association of Korea Journal*, 22(2), 61-93.
- Kim, H. (2020). Nominal modifiers in argumentative essays as discriminators for

- writing course placement decisions. *English Teaching*, 75(3), 3-24.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304.
- Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333-349.
- Lan, G., Liu, Q., & Staples, S. (2019). Grammatical complexity: ‘What does it mean’ and ‘so what’ for L2 writing classrooms? *Journal of Second Language Writing*, 46, 100673.
- Lan, G., Lucas, K., & Sun, Y. (2019). Does L2 writing proficiency influence noun phrase complexity? A case analysis of argumentative essays written by Chinese students in a first-year composition course. *System*, 85, 102116.
- Lan, G., & Sun, Y. (2019). A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course. *Journal of English for Academic Purposes*, 38, 14-24.
- Lan, G., Zhang, Q., Lucas, K., Sun, Y., & Gao, J. (2022). A corpus-based investigation on noun phrase complexity in L1 and L2 English writing. *English for Specific Purposes*, 67, 4-17.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12(4), 439-448.
- Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H. W. Seliger, & M. H. Long (Ed.), *Classroom-oriented research in second language acquisition* (pp. 287-304). Newbury House. [Study 2, cross-sectional data; and Study 3, longitudinal data.]
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619.
- Larsen-Freeman, D., & Celce-Murcia, M. (2016). *The grammar book: Form,*

- meaning and use for English language teachers*. National Geographic Learning with Heinle Cengage Learning.
- Larsen-Freeman, D., & Strom, V. (1977). The construction of a second language acquisition index of development. *Language Learning*, 27(1), 123-134.
- Lee, H. K. (2011). Investigating the applicability of the CEFR to a placement test for an English language program in Korea. *English Language and Linguistics*, 17(3), 29-60.
- Li, C., & Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.
- Li, Y. (2000). Linguistic characteristics of ESL writing in task-based e-mail activities. *System*, 28(2), 229-245.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36-62.
- Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16-27.
- McCarthy, M. (2016). Putting the CEFR to good use: Designing grammars based on learner-corpus evidence. *Language Teaching*, 49(1), 99-115.
- McHugh, M. (2013). The chi-square test of independence. *Biochemia Medica*, 23(2), 143-149.
- McNamara, T. F. (1990). Item response theory and the validation of an ESP test for health professionals. *Language Testing*, 7(1), 52-76.
- Monroe, J. H. (1975). Measuring and enhancing syntactic fluency in French. *The French Review*, 48(6), 1023-1031.
- Musgrave, J., & Parkinson, J. (2014). Getting to grips with noun groups. *Elt*

- Journal*, 68(2), 145-154.
- Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 2-13.
- Neff, J., Dafouz, E., Diez, M., Prieto, R., & Chaudron, C. (1998). Contrastive discourse analysis: Argumentative text in English and Spanish. Paper presented at the 24th Linguistics Symposium: Discourse across Languages and Cultures. University of Wisconsin–Milwaukee.
- Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading & Writing Quarterly*, 23(3), 287-309.
- Nichols, J. (1984). Functional theories of grammar. *Annual Review of Anthropology*, 13(1), 97-117.
- Nichols, J. (2009). Linguistic complexity: A comprehensive definition and survey. In G. Sampson, D. Gil, & P. Trudgill (Ed.), *Language complexity as an evolving variable* (pp. 110-125). Oxford University Press.
- Nihalani, N. K. (1981). The quest for the L2 index of development. *RELC Journal*, 12(2), 50-56.
- Norrby, C., & Håkansson, G. (2007). The interaction of complexity and grammatical processability: The case of Swedish as a foreign language. *International Review of Applied Linguistics*, 45, 45–68.
- Norris, J., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, 51(S1), 157-213.
- Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty, & M. H. Long (Ed.), *The handbook of second language acquisition* (pp. 716-761). Blackwell.
- Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578.

- Oh, S. Y. (2007). A corpus-based study of epistemic modality in Korean college students' writings in English. *English Teaching*, 62(2), 147-175.
- Oh, S. Y. (2014). Use of signaling nouns in published and Korean student academic writing in applied linguistics. *English Teaching*, 69(1), 147-172.
- Oh, S. Y., & Kang, S. (2013). The effect of English proficiency on Korean undergraduates' expression of epistemic modality in English argumentative writing. *Journal of Asia TEFL*, 10(4), 97-132.
- Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, 19(4), 218-233.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518.
- Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann, & B. Szmrecsanyi (Ed.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127-155). Mouton de Gruyter.
- Park, H. (2017). Noun phrases in Korean students' written English. *The Journal of Linguistics Science*, 81, 109-131.
- Parkinson, J. (2013). Adopting academic values: Student use of that-complement clauses in academic writing. *System*, 41(2), 428-442.
- Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14, 48-59.
- Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly*, 14(1), 61-69.
- Perkins, K., & Homburg, T. (1980). Three different statistical analyses of objective measures of attained writing proficiency. In R. Silverstein (Ed.),

- Occasional papers in linguistics, No. 6* (pp. 306-317). Southern Illinois University.
- Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System*, 38(3), 444-456.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Rahimpour, M. (2007). Task complexity and variation in L2 learners' oral discourse. *The University of Queensland Working Papers in Linguistics*, 1, 1-9.
- Reilly, J., Baruch, E., Jisa, H., & Berman, R. (2002). Propositional attitudes in written and spoken language. *Written Language & Literacy*, 5(2), 183-218.
- Reppen, R. (2001). Writing development among elementary students: Corpus-based perspectives. In R. Simpson, & J. Swales (Ed.), *Corpus linguistics in North America: Selections from the 1999 symposium* (pp. 211–225). University of Michigan Press.
- Révész, A., Kourтали, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, 67(1), 208-241.
- Rhee, S. C., & Jung, C. K. (2014). Compilation of the Yonsei English Learner Corpus (YELC) 2011 and its use for understanding current usage of English by Korean pre-university students. *The Journal of the Korea Contents Association*, 14(11), 1019-1029.
- Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing*, 23(4), 497-519.
- Ruiz-Funes, M. (2014). Task complexity and linguistic performance in advanced college-level foreign language writing. In H. Byrnes, & R. Manchon (Ed.), *Task-based language learning: Insights from and for L2 writing* (pp. 163–192). John Benjamins.

- Sackett, P. R., Laczo, R. M., & Arvey, R. D. (2002). The effects of range restriction on estimates of criterion interrater reliability: Implications for validation research. *Personnel Psychology, 55*(4), 807-825.
- Scott, M. (2004). *Oxford WordSmith Tools Version 4.0*. Oxford University Press.
- Sharma, A. (1980). Syntactic maturity: Assessing writing proficiency in a second language. In R. Silverstein (Ed.), *Occasional papers in linguistics, No. 6* (pp. 318– 325). Southern Illinois University.
- Sharpe, D. (2015). Chi-square test is statistically significant: Now what? *Practical Assessment, Research, and Evaluation, 20*(1), 8.
- Sharwood Smith, M., & Kellerman, E. (1986). Crosslinguistic influence in second language acquisition: An introduction. In E. Kellerman, & M. Sharwood Smith (Eds.), *Crosslinguistic influence in second language acquisition* (pp. 1-9). Pergamon Press.
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*(1), 103-110.
- Skehan, P. (1989). *Individual differences in second-language learning*. Edward Arnold.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510-532.
- Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexicogrammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing, 32*, 17-35.
- Storch, N., & Wigglesworth, G. (2007). Writing tasks: The effects of collaboration. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 157–177). Multilingual Matters.
- Taguchi, N., Crawford, W., & Wetzell, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college

- composition program. *TESOL Quarterly*, 47(2), 420-430.
- Tedick, D. J. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. *English for Specific Purposes*, 9(2), 123-143.
- Uccelli, P., Dobbs, C. L., & Scott, J. (2013). Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication*, 30(1), 36-62.
- Van Rooy, B., & Schäfer, L. (2002). The effect of learner errors on POS tag errors during automatic POS tagging. *Southern African Linguistics and Applied Language Studies*, 20(4), 325-335.
- Warchał, K. (2010). Moulding interpersonal relations through conditional clauses: Consensus-building strategies in written academic discourse. *Journal of English for Academic Purposes*, 9(2), 140-150.
- Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal*, 84(2), 171-184.
- White, P. R., & Sano, M. (2006). Dialogistic positions and anticipated audiences – A framework for stylistic comparisons. In K. Aijmer, & A. Simon-Vandenberg (Ed.), *Pragmatic markers in contrast* (pp. 189-214). Elsevier.
- Wingate, U. (2012). ‘Argument!’ helping students understand what essay writing is about. *Journal of English for Academic Purposes*, 11(2), 145-154.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity* (Technical Report No. 17). University of Hawaii.
- Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53-67.
- Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic,

proficiency, and construct multidimensionality. *System*, 66, 130-141.

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, 51(2), 275-301.

Ziegler, H. (1991). On translating "The Sunday Drive". *The Review of Contemporary Fiction*, 11, 137-149.

APPENDICES

APPENDIX 1	162
APPENDIX 2	163
APPENDIX 3	164

Appendix 1. A Sample Tagged Text

< Mid-rated essay, file no. 2170 >

I_PPIS1 think_VV0 that_CST drivers_NN2 of_IO automobiles_NN2 should_VM not_XX be_VBI allowed_VVN to_TO use_VVI cellular_JJ phones_NN2 while_CS driving_VVG . . .
these_DD2 days_NNT2 , , many_DA2 car_NN1 accidents_NN2 happen_VV0 because_II21 of_II22 drivers_NN2 ' _GE mistakes_NN2 . . .
one_MC1 of_IO their_APPGE mistakes_NN2 is_VBZ to_TO use_VVI cellular_JJ phones_NN2 while_CS driving_VVG . . .
using_VVG cellular_JJ phones_NN2 while_CS driving_VVG is_VBZ very_RG dangerous_JJ . . .
because_CS that_DD1 behavior_NN1 takes_VVZ away_RL driver_NN1 's_GE attention_NN1 from_II driving_VVG to_II using_VVG cellular_JJ phones_NN2 . . .
so_RR they_PPHS2 ca_VM n't_XX concentrate_VVI with_IW driving_JJ . . .
then_RT , , that_DD1 makes_VVZ the_AT situation_NN1 which_DDQ is_VBZ easy_JJ to_TO generate_VVI car_NN1 accident_NN1 . . .
also_RR , , I_PPIS1 think_VV0 using_VVG cellular_JJ phones_NN2 while_CS driving_VVG can_VM kill_VVI innocent_JJ people_NN . . .
if_CS car_NN1 accidents_NN2 happend_VV0 , , both_RR drivers_NN2 and_CC pedestrians_NN2 are_VBR dangered_NNU so_RR , , innocent_JJ people_NN are_VBR hurt_VVN because_II21 of_II22 one_MC1 person_NN1 's_GE mistake_NN1 . . .
many_DA2 drivers_NN2 do_VD0 n't_XX think_VVI that_CST their_APPGE mistakes_NN2 can_VM take_VVI away_RL other_JJ 's_GE life_NN1 . . .
so_RR they_PPHS2 need_VV0 to_TO think_VVI about_II their_APPGE responsibility_NN1 for_IF other_JJ people_NN . . .
because_II21 of_II22 these_DD2 reason_NN1 , , I_PPIS1 think_VV0 that_CST drivers_NN2 of_IO automobiles_NN2 should_VM not_XX be_VBI allowed_VVN to_TO use_VVI cellular_JJ phones_NN2 while_CS driving_VVG . . .

Appendix 2. Concordance Lines of Attributive Adjectives

< High-rated essay, file no. 358 >

```
Line 1: Though_CS there_EX can_VM be_VBI many_DA2 opinions_NN2 about_II this_DD1 ,_, I_PPIS1 firmly_RR believe_VV0 that_CST not_XX all_DB Koreans_NN2 have_VH0 to_TO complete_VVI military_JJ service_NN1 . . .
Line 3: First_MD ,_, population_NN1 is_VBZ not_XX what_DDQ the_AT army_NN1 needs_NN2 and_CC second_NNT1 ,_, there_EX are_VBR some_DD talented_JJ people_NN whose_DDQGE talents_NN2 should_VM not_XX be_VBI wasted_V
Line 4: First_MD and_CC foremost_JJT ,_, the_AT main_JJ part_NN1 in_II today_RT 's_GE war_NN1 is_VBZ not_XX the_AT number_NN1 of_IO soldiers_NN2 . . .
Line 5: From_II Iraq_NP1 war_NN1 of_IO United_NP1 States_NP1 to_II WWII_NN1 in_II which_DDQ the_AT nuclear_JJ bomb_NN1 was_VBDZ introduced_VVN ,_, the_AT development_NN1 of_IO massive-destructing_JJ weapons_NN2
Line 6: From_II Iraq_NP1 war_NN1 of_IO United_NP1 States_NP1 to_II WWII_NN1 in_II which_DDQ the_AT nuclear_JJ bomb_NN1 was_VBDZ introduced_VVN ,_, the_AT development_NN1 of_IO massive-destructing_JJ weapons_NN2
Line 8: Moreover_RR ,_, forcing_VVG some_DD talented_JJ people_NN to_TO fulfill_VVI military_JJ service_NN1 is_VBZ wasting_VVG their_APPGE talents_NN2 . . .
Line 8: Moreover_RR ,_, forcing_VVG some_DD talented_JJ people_NN to_TO fulfill_VVI military_JJ service_NN1 is_VBZ wasting_VVG their_APPGE talents_NN2 . . .
Line 9: Talents_NN2 of_IO entertainers_NN2 and_CC sports_NN2 players_NN2 can_VM be_VBI used_VVN to_TO upgrade_VVI national_JJ brand_NN1 by_II winning_VVG plays_NN2 during_II World_NN1 Cup_NN1 or_CC creating_VVG
Line 9: Talents_NN2 of_IO entertainers_NN2 and_CC sports_NN2 players_NN2 can_VM be_VBI used_VVN to_TO upgrade_VVI national_JJ brand_NN1 by_II winning_VVG plays_NN2 during_II World_NN1 Cup_NN1 or_CC creating_VVG
Line 9: Talents_NN2 of_IO entertainers_NN2 and_CC sports_NN2 players_NN2 can_VM be_VBI used_VVN to_TO upgrade_VVI national_JJ brand_NN1 by_II winning_VVG plays_NN2 during_II World_NN1 Cup_NN1 or_CC creating_VVG
Line 11: D_ZZ1 course_NN1 are_VBR in_II their_APPGE golden_JJ age_NN1 of_IO their_APPGE career_NN1 . . .
Line 12: Their_APPGE talents_NN2 should_VM be_VBI used_VVN for_IF the_AT society_NN1 by_II serving_VVG alternative_JJ military_JJ service_NN1 such_II21 as_II22 working_VVG for_IF industrial_JJ development_NN1 an
Line 12: Their_APPGE talents_NN2 should_VM be_VBI used_VVN for_IF the_AT society_NN1 by_II serving_VVG alternative_JJ military_JJ service_NN1 such_II21 as_II22 working_VVG for_IF industrial_JJ development_NN1 an
Line 13: To_TO wrap_VVI it_PPH1 up_RP ,_, I_PPIS1 argue_VV0 that_CST forcing_VVG all_DB Korean_JJ to_TO complete_VVI military_JJ service_NN1 is_VBZ not_XX necessary_JJ for_IF the_AT reasons_NN2 written_VVN above
Line 14: However_RR ,_, I_PPIS1 am_VBM not_XX insisting_VVG that_DD1 military_JJ service_NN1 itself_PPX1 is_VBZ unnecessary_JJ . . .
Line 15: Providing_VVG people_NN with_IW diverse_JJ ways_NN2 to_TO serve_VVI for_IF the_AT society_NN1 insted_VVD of_IO just_RR the_AT military_NN1@ is_VBZ my_APPGE main_JJ point_NN1 . . .
Line 15: Providing_VVG people_NN with_IW diverse_JJ ways_NN2 to_TO serve_VVI for_IF the_AT society_NN1 insted_VVD of_IO just_RR the_AT military_NN1@ is_VBZ my_APPGE main_JJ point_NN1 . . .
```

Note. JJ = General adjective, NN = Common noun, neutral for number, NN1 = Singular common noun, NN2 = Plural common noun. The highlighted parts indicate attributive adjective-noun sequences.

Appendix 3. Observed and Expected Count Output from SPSS

		Low	Intermediate	High	Total (row)
FAC	Obs	88	221	221	530
	Exp	65.1	200.2	264.8	
	Adj	3.2	2.0	-4.0	
VAC	Obs	53	130	193	376
	Exp	46.2	142.0	187.8	
	Adj	1.1	-1.3	0.6	
NC	Obs	5	35	50	90
	Exp	11.0	34.0	45.0	
	Adj	-2.0	0.2	1.1	
TRC	Obs	19	33	55	107
	Exp	13.1	40.4	53.5	
	Adj	1.7	-1.5	0.3	
WRC	Obs	19	123	103	245
	Exp	30.1	92.5	122.4	
	Adj	-2.2	4.1	-2.5	
PA	Obs	404	1191	1531	3126
	Exp	383.7	1180.6	1561.6	
	Adj	1.6	0.6	-1.6	
PN	Obs	43	126	196	365
	Exp	44.8	137.9	182.3	
	Adj	-0.3	-1.3	1.5	
PP <i>of</i>	Obs	42	170	336	548
	Exp	67.3	207.0	273.8	
	Adj	-3.5	-3.4	5.6	
PP	Obs	32	140	184	356
	Exp	43.7	134.5	177.8	
	Adj	-2.0	0.6	0.7	
Total (column)		705	2169	2869	5743

Note. FAC = Finite adverbial clauses, VAC = Verb/Adjective complement clauses, NC = Noun complement clauses, TRC = *That* relative clauses, WRC=WH relative clauses, PA = Premodifying adjectives, PN = Premodifying nouns, PP *of*= Prepositional phrases *of*, PP = Prepositional phrases other than *of*. Obs means the observed count; Exp means the expected count; and Adj means the adjusted residuals. Exp of a cell is computed as (raw total * column total) / grand total (McHugh, 2013).

국 문 초 록

한국 대학생들의 논증적 에세이에 나타난
절과 구 복잡성의 발달

서 나 래
외국어교육과 영어전공
서울대학교 대학원

영어 글쓰기 발달에 관한 연구들은 문법적 복잡성(grammatical complexity)을 학습자의 능숙도를 구별하는 중요한 지표로 인식하고 있다. 초기 연구들은 주로 절 복잡성(clausal complexity)에 기반해 문법적 복잡성을 측정하였지만, 최근 연구들은 구 복잡성(phrasal complexity)에 초점을 두고 있다. 이러한 변화는 절 복잡성이 일상 대화가 가진 특징으로 글쓰기의 초기 발달 단계를 나타내는 반면, 구 복잡성, 특히 명사구의 복잡성은 학문적 글(academic writing)이 가진 복잡성의 전형으로써 높은 수준의 발달 단계를 나타낸다는 인식에 기반하고 있다. 하지만 일부 연구들은 명사구의 복잡성이 글쓰기 능숙도와 큰 관련이 없다는 상반된 결과를 보이고 있는데, 이는 대부분의 연구들이 학습자 모국어가 문법적 복잡성에 미치는 영향을 고려하지 않고 다양한 모국어를 가진 학습자들에 의해 만들어진 코퍼스를 사용했기 때문일 수 있다. 이에 본 연구는 한국인 대학생들이 작성한 글을 분석하여 절과 구의 복잡성이 글쓰기 능숙도와 연관성이 있는지 살펴보고, 그러한

연관성에 크게 기여한 복잡성 특징들을 바탕으로 문법적 복잡성의 발달 패턴을 추정하고자 하였다. 또한 학생들의 글을 질적으로 분석하여, 특정 복잡성 특징을 구현할 때 자주 쓰이는 어휘와 오류 빈도 및 유형을 파악함으로써 능숙도 집단 간의 차이를 더 자세히 묘사하고자 하였다.

본 연구에 사용된 코퍼스는 연세 영어 학습자 코퍼스(Yonsei English Learner Corpus, YELC 2011)에서 추출한 234개의 논증적 에세이로 구성되어 있으며, 이는 CEFR에 기반하여 초급, 중급, 고급의 글쓰기 능숙도를 나타내는 세 개의 하위 코퍼스로 구분되었다. 품사 태깅된 코퍼스를 바탕으로 정규표현식(regular expressions)을 사용하여, Biber et al. (2011)이 제안한 발달단계에 있는 9개의 절 복잡성 특징과 8개의 구 복잡성 특징을 추출하여 각각의 빈도를 계산하였다.

피어슨 카이제곱검정(a Pearson Chi-square test) 결과, 글쓰기 능숙도가 절과 구의 복잡성과 유의한 연관성이 있다는 결론이 도출되었다. 사후검정으로 잔차 분석(a residual analysis)을 수행한 결과, 특히 5개 복잡성 특징이 이러한 연관성에 크게 기여했음이 밝혀졌다. 주목할 만한 발견은 각 능숙도 집단의 주요 복잡성 특징이 Biber et al. (2011)이 제안한 발달단계와 일치하며 따라서 한국인 대학생의 발달 패턴이 두 개의 매개변수, 즉 (1) 구조적 형태와 (2) 통사적 기능에 의해 설명될 수 있다는 점이다. 즉, 한국 대학생들의 문법적 복잡성은 (i) 절의 구성 성분으로 기능하는 정형 종속절(finite dependent clauses functioning as clause constituents)인 부사절의 빈번한 사용에서 (ii) 명사구의 구성 성분으로 기능하는 정형 종속절(finite

clause types function as NP constituents)인 WH 관계절에 대한 의존을 거쳐 (iii) 명사구의 구성 성분으로 기능하는 종속구(dependent phrasal structures functioning as noun phrase constituents)인 *of* 전치사구에 대한 선호로 발달하는 것으로 나타났다.

예상과 달리, 명사의 선수식어(premodifier)로 사용되는 형용사 및 명사의 빈도는 글쓰기 능숙도와 큰 연관성이 없는 것으로 나타났다. 이에 관해 학생들의 글을 질적 분석한 결과, 첫째, 초급 수준의 글은 쓰기 지시문(writing prompts)에 제시된 형용사+명사 조합을 반복적으로 사용하는 경향을 보였다. 둘째, 명사+명사 구조와 관련한 오류가 능숙도가 높아질수록 현저히 낮아지는 경향을 보였다. 마지막으로, 보어절(complement clauses)과 관련해서는 모든 능숙도 수준의 학생들이 매우 한정적인 종류의 통제 명사(controlling nouns)를 사용했으며, 학문적인 글 보다는 일상 대화에서 쓰이는 통제 동사(controlling verbs)를 사용하였다.

이러한 연구 결과는 크게 세가지 교육적 함의를 시사한다. 첫째, 경험적으로 도출된 문법적 복잡성의 발달 단계를 상세한 평가 척도 설명자(rating scale descriptors) 개발과 보다 맞춤형 된 수업 설계를 위해 활용해야 한다. 둘째, 학문적인 글에서 보어절과 함께 자주 사용되는 통제 명사 및 동사에 대한 교실 수업을 통해, 학습자들이 문법적 구조를 학문적인 어휘로 실현할 수 있도록 해야 한다. 마지막으로, 특히 명사를 선수식하는 명사 및 관계대명사절의 사용에 있어 학습자의 글에서 자주 발견되는 오류를 시정함으로써, 문법 구조 사용에 대한 정확성을 향상시켜야 한다.

주요어: 문법적 복잡성, 학문적 글쓰기, 영어 글쓰기 발달, 절 복잡성, 구 복잡성

학 번: 2021-20090