# NOUN PHRASE COMPLEXITY, ACADEMIC LEVEL, AND FIRST- AND SECOND-ENGLISH LANGUAGE BACKGROUND IN ACADEMIC WRITING

by

**Ge Lan**

**A Dissertation**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Doctor of Philosophy**


Department of English

West Lafayette, Indiana

May 2020

**THE PURDUE UNIVERSITY GRADUATE SCHOOL**

**STATEMENT OF COMMITTEE APPROVAL**

**Dr. April Ginther, Chair**

Department of English

**Dr. Bradly Dilger**

Department of English

**Dr. Elaine Francis**

Department of English

**Dr. Shelley Staples**

Department of English, University of Arizona

**Approved by:**

Dr. Manushag Powell

*To my family*

*and*

*To those who have inspired me*

# ACKNOWLEDGMENTS

From May 2018 to April 2020, I have spent the past two years to complete my dissertation. Having gone through these two years, I have grown from a student who knows little about the field of second language studies to a mature emerging scholar who is able to publish research work. I would like to express my gratitude to many people, because without their supports, I would not have been able to complete my dissertation. Sincerely, I would like to thank the following people:

- Dr. April Ginther, my academic advisor and the committee chair, for her guidance on my dissertation work, her mentoring on my professional development, and her suggestions on possible research directions that I can follow in the future;
- Dr. Bradley Dilger, my committee member, for all research opportunities and academic resources he has provided me in the Corpus and Repository of Writing, which effectively help me develop my corpus and computational skills;
- Dr. Shelley Staples, my committee member (from University of Arizona), for her generosity of helping me on my dissertation and relevant research, and her mentoring on my professional development as an emerging scholar in corpus linguistics;
- Dr. Elaine Francis, my committee member, for her feedback on my dissertation and her suggestions on a research direction that I can follow in the future;
- Mr. Mark Fullmer, the website developer of Corpus and Repository of Writing, for his help on program debugging for my dissertation;
- My peers in Second Language Studies, for their helps on data processing for my dissertation, including Dr. Yachao Sun, Dr. Jie Gao, Qiusi Zhang, and Kyle Lucas.

Also, I would like to thank my parents for their support from the beginning to the end of my time in the United States. I came to the United States in January 2011 as an undergraduate exchange student, and I have then stayed in this country until now, April 2020. It is hard to believe that I have been absent from my parents' life for almost the past ten years. Their support on all my decisions in the past ten years has made me who I am now. At the end, I would like to thank my fiancée, Yi Li (Eva). We were engaged in June 2016, and she has been waiting to marry me for years. People often say that a successful man has a strong woman behind him. I am not sure if I am a successful man but thanks for backing me up these years, Yi Li. You are always the strong woman behind me.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

Since the 1990s, grammatical complexity is a topic that has received considerable attention in various fields of applied linguistics, such as English for academic purposes, second language acquisition, language testing, and second language writing (Bulté & Housen, 2012). Many scholars in applied linguistics have recently argued that grammatical complexity has primarily been represented by clausal features (e.g., subordinate clauses), and it is important to study grammatical complexity as a multidimensional construct based on both clausal features and phrasal features (Biber, Gray & Poonpon, 2011; Norris & Ortega, 2009). Thus, this dissertation is a corpus-based investigation on how the use of noun phrases is influenced by two situational characteristics of a university context: academic level and first- and second-English language background.

I built my corpus by extracting 200 essays from British Academic Written English Corpus, which represents academic writing of (1) undergraduate and graduate students and (2) L1 and L2 students. Noun phrase complexity was then operationalized to the 11 noun modifiers proposed in the hypothesized developmental index of writing complexity features in Biber, Gray and Poonpon (2011). The 11 noun modifiers were extracted from the corpus and counted for statistical analysis via a set of Python programs. With a Chi-square test followed by a residual analysis, I found that both academic level and first- and second-English language background influenced noun phrase complexity but in distinct ways. The influence of academic level is primarily associated with three phrasal modifiers (i.e., attributive adjectives, premodifying nouns, and appositive NPs) and two clausal modifiers (i.e., relative clauses and noun complement clauses). The undergraduate corpus includes more of the two clausal modifiers, whereas the graduate corpus has more of the three phrasal modifiers. This suggests that, in these

200 essays, graduate students tend to build more compressed NPs than undergraduate students. However, the influence of first- and second-English language background derives from a much broader range of noun modifiers, including eight noun modifiers (e.g., attributive adjectives, relative clauses, infinitive clauses). More diverse NP patterns with different noun modifiers are in the L1 corpus than in the L2 corpus. Surprisingly, the L2 corpus has more phrasal noun modifiers (i.e., attributive adjectives, premodifying nouns), which has been argued to indicate advanced levels of academic writing. A qualitative analysis on selected essays reveals that some cases of attributive adjectives and premodifying nouns are repeatedly used by L2 students to help content development in their writing. Overall, this dissertation adds an additional piece of evidence on the importance of noun phrase complexity in writing research.

# CHAPTER 1. **INTRODUCTION**

The development of academic writing skills has received substantially educational concerns in universities, because writing skills are considered an important way to evaluate knowledge in academic communities (Staples, Egbert, Biber, & Gray, 2016). Since the 1990s, researchers in applied linguistics have worked extensively on assessing academic writing from university students, especially second language (L2) students. The CAF model (i.e., complexity, accuracy, and fluency) was originally used in first language (L1) studies and was then introduced in different fields of L2 studies, for example, second language acquisition (SLA), writing assessment, English for Academic Purposes (EAP), and L2 writing (Housen & Kuiken, 2009; Norris & Ortega, 2009). An important construct, *complexity*, has had a long tradition in L2 writing because researchers have gathered empirical support showing a close relationship between (a) grammatical complexity and (b) L2 writing development and writing proficiency (Biber, Gray, & Poonpon, 2011).

A particular type of grammatical complexity has recently received extensive attention among researchers in L2 writing in the 2010s, namely, noun phrase (NP) complexity (Parkinson & Musgrave, 2014). The reasons behind this research trend can be summarized as the following two points. First, previous research on grammatical complexity has mainly focused on clausal complexity constructed by subordinate structures, and researchers argue that phrasal complexity should be integrated into grammatical analysis to make grammatical complexity comprehensively represented (e.g., Biber et al., 2011; Bulté & Housen, 2012). Second, NPs are a frequent characteristic of advanced academic writing (e.g., graduate writing), and construction of NPs is argued to be indicative for the development of academic writing skills (Biber et al., 2011; Parkinson & Musgrave, 2014; Staples et al., 2016). As a result, some recent empirical studies

have been conducted to explore NP complexity among different groups of writers in universities (e.g., EAP students, L2 students, graduate students).

Having said this, grammatical complexity has been inconsistently defined and operationalized across even within empirical studies (Bulté & Housen, 2012), which affects how NP complexity has been studied in writing research (Lan, Lucas & Sun, 2019). In addition, although grammatical complexity has been substantially investigated in academic writing over the past 20 years, only a few recent studies focus on NP complexity in particular. In this dissertation, I follow the recent research trend of investigating NP complexity, an underrepresented component of grammatical complexity in EAP and L2 writing research. The purpose of this dissertation is to examine how NP complexity is influenced by two variables closely associated with academic writing: academic level and first- and second-English language background (i.e., L1- and L2-English language background). I first review changes regarding discussion of grammatical complexity in L2 writing and argue for the use of register-based perspective to study NP complexity in my dissertation. Then, I quantitatively analyze the relationships between NP complexity and the two variables (i.e., academic level, L1- and L2-English language background), respectively. The quantitative findings are interpreted based on a qualitative analysis of specific writing examples from different group of writers (e.g., L2 undergraduate students). Although this dissertation tends to be exploratory and descriptive, some research and pedagogical implications are provided at the end to outline future's research path on the topic of NPs and to share suggestion on NP instruction in writing classroom.

# CHAPTER 2. GRAMMATICAL COMPLEXITY

## 2.1. Review of Grammatical Complexity

Since the 1990s, grammatical complexity has received substantial research attention in various areas of L2 studies, such as second language acquisition (SLA), English for Academic Purposes (EAP), language testing, and L2 writing (Bulté & Housen, 2012). Wolfe-Quintero et al. (1998) completed the first comprehensive review of grammatical complexity along with two other important constructs (i.e., accuracy and fluency), which make this construct a central issue in L2 writing. From then on, a wealth of cross-sectional studies (e.g., Biber et al., 2011; Casal & Lee, 2019; Lan, Lucas, & Sun, 2019; Lu, 2011; Parkinson & Musgrave, 2014; Staples, Egbert, Biber, & Gray, 2016; Wang & Beckett, 2017) and longitudinal studies (e.g., Crossley & McNamara, 2014; Mazgutova & Kormos, 2015; Ravid & Berman, 2010; Spoelman & Verspoor, 2010; Vyatkina, 2013; Yoon & Polio, 2017) have been conducted to explore grammatical complexity.

Although grammatical complexity has been substantially investigated in writing research, this construct remains inconsistently conceptualized and operationalized by scholars. To address the problem of the conceptual and operational inconsistencies, it is critical to review the changes of how grammatical complexity has been discussed among scholars in writing research. As Wolfe-Quintero et al. (1998) has been considered a milestone work on complexity in L2 writing, I used this publication as a starting point to review on the discussion of grammatical complexity in the recent 20 years (1998 – 2018). Also, it is important to mention that there are different perspectives to study grammatical complexity in writing. For instance, studies from the psycholinguistic perspective have primarily focused on the relationship between grammatical complexity and efficiency of language processing (e.g., Gibson, 1998; Hawkins, 2004), whereas

studies from the applied linguistic perspective have often focused on how grammatical forms and/or measures can be applied to analyze writing development and writing proficiency (e.g., Biber et al., 2011; Norris & Ortega, 2009; Wolfe-Quintero et al., 1998). In this chapter, my review of grammatical complexity is based on existing literature regarding L2 writing in applied linguistics.

This chapter discusses the changes in the conceptualization and operationalization of grammatical complexity in L2 writing over the past 20 years (1998 – 2018) to build a solid theoretical foundation for this dissertation. Two specific sections are included in this chapter, namely, a conceptual review of grammatical complexity and an operational review of grammatical complexity. As "[n]oun phrases vary greatly both in structure and complexity, and they can have a wide range of syntactic roles" (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p. 230), this review helps me answer two important questions: (1) why is NP complexity important to study? (2) How can NP complexity be effectively studied?

## 2.2. A Conceptual Review of Grammatical Complexity

From the conceptual perspective, grammatical complexity can be viewed as an abstract property of a system (Bulté & Housen, 2012). In the past two decades (1998 – 2018), discussions of grammatical complexity in writing research reveal two noteworthy problems: first, there are inconsistencies regarding how this construct is defined; second, there is a challenge with representing this construct.

### *2.2.1. Inconsistency on Definition of Grammatical Complexity*

Since the 1990s, grammatical complexity has remained inconsistently defined in L2 studies. Bulté and Housen (2012) recently emphasized this theoretical problem: "there is no

consensus in the literature on the definition of complexity, and no consistency as to how it has been operationalized across (and sometimes even within) studies" (p.43). In L2 writing (and applied linguistics in general), grammatical complexity has been conceptualized in two ways based on whether contexts should or should not be taken into consideration. Bulté and Housen (2012) mention an absolute approach without regard to context: complexity can be defined in an objective term as the number of grammatical features and as the connections among these features. In register studies, Biber et al. (2011) claim that grammatical features should be analyzed across situations of language use (which they refer to as *registers*), because grammatical features are often used to meet specific grammatical functions in particular situation. I refer Biber et al.'s (2011) claim as "register approach" in this chapter. So far, these two types of approaches, absolute and register, co-exist in L2 writing research.

In terms of the absolute approach of defining grammatical complexity, some existing definitions are provided to present how grammatical complexity has developed conceptually. Wolfe-Quintero et al. (1998) define grammatical complexity in L2 writing as "a wide variety of both basic and sophisticated structures" (p. 69). This definition should be carefully examined for two reasons. First, "sophisticated structures" need to be clarified. Second, it is debatable how "basic structures" represent grammatical complexity. For example, a simple NP (e.g., a book) may not suggest any complexity with regard to language use. Then, Rimmer (2006) argued that grammatical complexity is based on the cumulative effect of a number of factors, such as sentence length, embedding, ellipsis, and markedness. All these factors contribute to grammatical complexity in writing. Although this definition is inclusive, Rimmer did not mention how these factors can be measured and/or applied in a practical context.

Recently, Lu (2011) proposed an effective description of grammatical complexity in the context of L2 writing: "Syntactic complexity is evident in second language (L2) writing in terms of syntactic variation and sophistication, or, more specifically, the range of syntactic structures that are produced and the degree of sophistication of such structures" (p. 36 - 37). Lu mentions "sophistication" of grammatical structures, but no measures directly related to "sophistication" were integrated into his study. Similar to Wolfe-Quintero et al. (1998), "sophistication" also needs to be clarified. Despite the difficulty of explicit measures indicative of "sophistication", Lu proposes two major senses of grammatical complexity, variation and sophistication[1], "or, more specifically, the range of syntactic structures that are produced and the degree of sophistication of such structures" (p. 36).

These two senses of grammatical complexity are further supported by Bulté and Housen's (2012) Taxonomy of L2 Complexity. In their taxonomy, a specific category of complexity is included: linguistic complexity, along with complexity in two other categories (i.e., discourse-interactional complexity and propositional complexity). For linguistic complexity, Bulté and Housen argue that grammatical complexity should encompass two components, which are consistent with the two senses of complexity in Lu (2011):

1. *systemic complexity*, referring to the 'breadth' of grammatical features an L2 learner can produce (i.e., grammatical variation)
2. s*tructural complexity*, representing the 'depth' of grammatical construction an L2 learner can compose (i.e., grammatical sophistication).

In recent publications, scholars in L2 studies (and applied linguistics in general) agree on these two components of grammatical complexity. For example, Crossley and McNamara (2014)

[1]The term "syntactic complexity" is used in Lu (2011). In the literatures of linguistic complexity in L2 writing, "syntactic complexity" and "grammatical complexity" are used interchangeably (Berlage, 2014).

claimed that grammatical complexity should cover "the sophistication of syntactic forms produced by a speaker or writer and the range or variety of syntactic forms produced" (p.68). It is important to mention that Bulté and Housen define "[grammatical] complexity in objective, quantitative terms as the number of discrete components that a language feature or a language system consists of, and as the number of connections between the different components" (Bulté & Housen, 2012, p. 24). This suggests that grammatical complexity is an independent construct, which can be objectively measured across different contexts.

In addition to Bulté and Housen's definition, scholars in register studies have recently come to believe that grammatical complexity should be considered with respect to contexts of language use, which is referred as the register approach in this dissertation. Register refers to a variety of language associated with a situation of use (Biber & Conrad, 2009). Biber et al. (2011) propose a form-function matrix to analyze grammatical complexity from the register perspective, and I created a figure to visualize the matrix (see Figure 1 below). The matrix has two components (i.e., grammatical form and grammatical function). *Grammatical form* refers to specific lexico-grammatical features, and *grammatical function* refers to the syntactic functions that the features play in a certain register. *Grammatical form* includes finite dependent clauses, nonfinite dependent clauses, and dependent phrases. *Grammatical function* covers adverbials, complements, and noun modifiers. Each grammatical form is associated with certain grammatical functions.

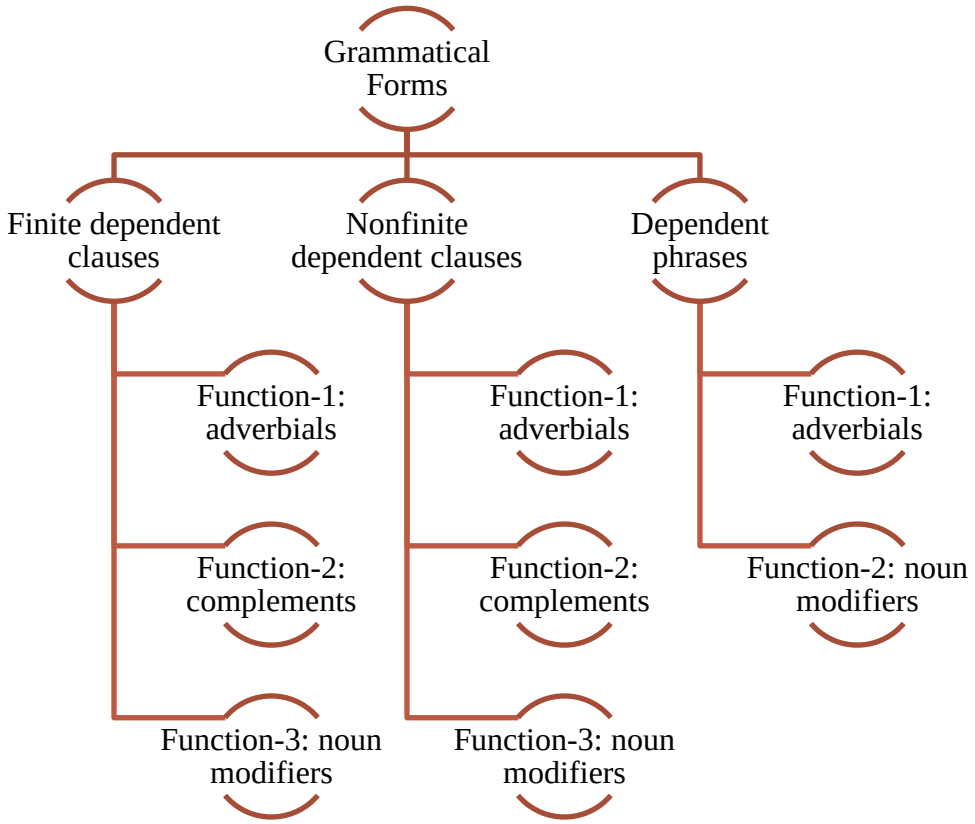



Figure 1. This figure demonstrates Biber, Gray and Poonpon (2011)'s Form and Function Matrix of Grammatical Complexity

The core spirit of the register approach is that grammatical forms are functional. Without taking register into consideration, the use of certain grammatical forms cannot be appropriately interpreted. For example, subordinate clauses are syntactically more complex than attributive adjectives. The use of subordinate clauses often suggests the development of grammatical complexity in L2 writing of ESL students, whereas the use of attributive adjectives does not (Larsen-Freeman, 1978). However, in academic writing, attributive adjectives are often used to build compressed NPs (i.e., nouns with phrasal modifiers) as in *an inconsistent research construct in the empirical studies*. Attributive adjectives enable the construction of compressed NPs that pack intensive information in a compact sentence structure, which is a characteristic of

advanced writing (Biber et al., 2011; Staples et al., 2016). In research articles, the use of the syntactically simpler feature (i.e., attributive adjectives) actually indicates the development of grammatical complexity to a more advanced stage. Therefore, the register approach of grammatical complexity helps us avoid potential misinterpretations of grammatical forms.

Overall, the definition of grammatical complexity has remained inconsistent in writing research. So far, this construct has been defined through the absolute approach (Bulté & Housen, 2012) and through the register approach (Biber et al., 2011). These two approaches are distinct from each other on whether contexts of language use should be taken into consideration, and they co-exist in L2 writing and applied linguistics in general.

### *2.2.2. Challenge on Representation of Grammatical Complexity*

Scholars in different periods of time have selected different grammatical features to represent grammatical complexity in writing studies. The representation of grammatical complexity has been challenged for the dominant focus on clausal complexity, whereas phrasal complexity is absent in the representation. In the past two decades (1998 – 2018), the selection of grammatical features has shifted from clausal features to both clausal features and phrasal features. Argued to be a critical type of phrasal features, NPs should be added to supplement clausal features to represent grammatical complexity in writing research. The integration of phrasal complexity (primarily NP complexity) is important for studying writing development.

Bulté and Housen (2012) pointed out an important distinction between "grammatical features that can be observed" and "grammatical features that are chosen to be observed". In the taxonomy, they listed three grammatical categories that can be observed in L2 writing: (1) morphological features, i.e., inflectional and derivational morphemes; (2) lexical features, i.e., words and collocations; and (3) syntactic features, i.e., phrases, clauses, and sentences. Each

grammatical category incorporates a number of features. For instance, clauses also include a number of specific types, such as causative clauses, conditional clauses, adverbial clauses, adjectival clauses, and nominal clauses. Therefore, scholars can technically observe a large number of grammatical features, but they often select the variables that most clearly demonstrate their arguments about grammatical complexity.

In L2 writing, researchers have chosen different grammatical features to represent grammatical complexity in their research, which has led to the changes in the representation of grammatical complexity. In early literature, three specific types of linguistic units were frequently selected: T-units[2], clauses, and sentences (Hunt, 1965). Wolfe-Quintero et al. (1998) conducted a comprehensive review of grammatical complexity based on these variables. However, in addition to T-units, clauses, and sentences, Wolfe-Quintero et al. also identified two other categories: first, particular grammatical features (e.g., adjectival clauses and adverbial clauses) and second, "the presence of specific grammatical structures in relation to clauses [e.g., dependent clauses per clause], T-units [e.g., clauses per T-unit], or sentences [e.g., subordinate clauses per sentence]" (p. 69). See Wolfe-Quintero et al.'s (1998) summary of the particular grammatical features below.

Particular Features in Wolfe-Quintero et al. (1998)

- Reduced clauses (1), dependent clauses (3), passives (1), passive sentences (2), adverbial clauses (1), adjectival clauses (2), nominal clauses (1), prepositional phrases (1), attributive adjectives (1), transitional connectors (1), subordinate clauses (1), coordinate clauses (2).

*Note*. The number of studies that applied each grammatical structure are shown in parentheses.

[2] "T-unit" stands for minimal terminal unit, which refers to a structure with a main clause as well as all the subordinate clauses attached to it (Hunt, 1965).

The summary of the particular features shows that only two phrasal features (i.e., attributive adjectives and prepositional phrases) were analyzed, each in only a single study, while the other ten remained clausal features. However, the analysis of these particular features in relation to T-units, clauses, and sentences can also be considered primarily clause-based. Wolfe-Quintero et al. (1998) reveals that although two additional categories of features were added to be observed, the representation of grammatical complexity was still dominated by clauses in L2 writing.

In the 2000s, Norris and Ortega (2009) pointed out that the selected linguistic units were primarily based on clauses, especially subordinate clauses. They argued that it was important to study grammatical complexity as a multi-dimensional construct, not only including subordinate structures, but also coordinate structures and phrasal structures. In recent years, more researchers have come to agree with Norris and Ortega (2009) that clausal and phrasal features should be integrated to represent a broader picture of grammatical complexity (e.g., Biber et al., 2011; Bulté & Housen, 2012; Staples et al., 2016; Yang, Lu, & Weigle, 2015). Although scholars define grammatical complexity as an independent or dependent construct, there is a consensus on the integration of phrasal structures (primarily NPs) as an important subcomponent of grammatical complexity.

Biber et al. (2011) proposed capturing grammatical complexity with the form-function matrix for grammatical analysis presented in Figure 1. In their matrix, grammatical form includes clausal features (i.e., finite dependent clauses and nonfinite clauses[3]) and phrasal features (i.e., dependent phrases). Biber et al. (2011) also draw attention to a type of grammatical complexity that is common in academic writing: phrasal complexity. Phrasal complexity refers

[3] Nonfinite clauses are categorized as a type of clausal feature in this case.

to noun-headed phrases with phrasal modifiers (e.g., premodifying nouns and prepositional phrases). Bulté and Housen (2012) also include phrasal complexity to represent a broader picture of grammatical complexity in the Taxonomy of L2 Complexity, and I created a figure to visualize the representation of grammatical complexity in the taxonomy (see Figure 2 below). Bulté and Housen include syntactic complexity and morphological complexity as two components of complexity. Syntactic complexity covers three subcomponents: sentential, clausal, and phrasal complexity. It is also interesting to see that morphological complexity has been integrated in grammatical complexity. However, morphological complexity has rarely been observed in studies of L2 writing in English, because it tends to be "morphologically poor" compared to other languages (DeClerq & Housen, 2016).

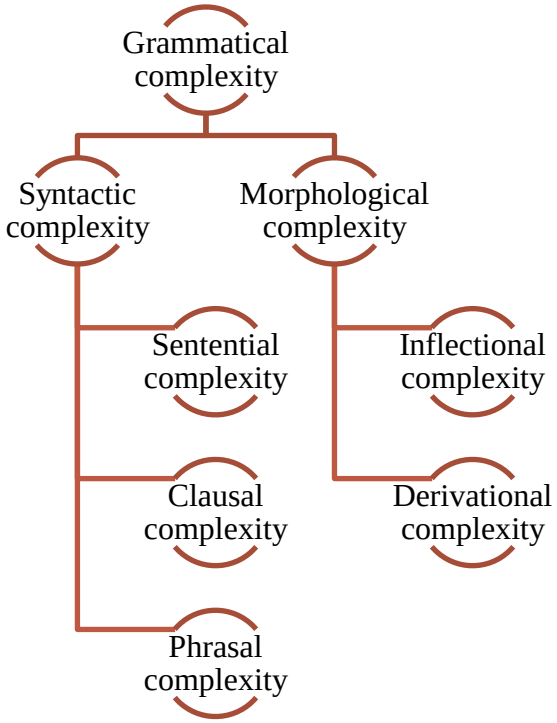



Figure 2. This figure presents the conceptualization of grammatical complexity in Bulté and Housen (2012)'s Taxonomy of L2 Writing Complexity

The integration of phrasal complexity into grammatical complexity is important for empirical studies that investigate the relationship between grammatical complexity and writing development. Scholars have summarized writing development as:

> [T]he stages generally progress from finite dependent 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, p.29 - 30).

In terms of writing development, phrasal features play an important role for NP construction at the advanced stage (Biber et al., 2011; Parkinson & Musgrave, 2014; Staples et al., 2016). As a result, grammatical complexity represented by clausal features can only capture writing development from the beginning to the intermediate stages but cannot capture writing development from the intermediate to the advanced stages. The integration of phrasal features (primarily NPs) into the representation of grammatical complexity helps researchers to explore writing development with a more comprehensive picture.

Overall, it has been a challenge for researchers to represent grammatical complexity in the past two decades (1998 – 2018). Echoing the claim that grammatical complexity needs to be studied as a multidimensional construct, phrasal features (primarily NPs) have been integrated into the representation of grammatical complexity. Grammatical complexity has been represented with a broader scope, which is important for studying writing development.

### 2.3. An Operational Review of Grammatical Complexity

From the operational perspective, grammatical complexity can be operationalized to measure the degree of complexity (Bulté & Housen, 2012). In L2 writing studies, researchers have operationalized grammatical complexity based on four major variables: *length*, *ratio*, *index*, and *frequency* (Lu, 2011; Norris & Ortega, 2009; Wolfe-Quintero et al., 1998). The application

of the four variables to operationalize grammatical complexity is associated with corresponding changes over the past two decades (1998 – 2018). The noteworthy changes are: (1) Length-based measures were applied as fluency measures in the 1990s but then as complexity measures in recent L2 writing research. (2) Ratio-based measures have been primarily used to capture subordinate structures instead of coordinate structures and phrasal structures. More measures need to be developed and/or applied for coordination and phrasal construction (e.g., NPs). (3) The index-based measures are difficult to calculate because they are often based on complex formulas. This makes the index variable rarely applied and discussed in L2 writing research. (4) Frequency-based measures have recently been operationalized to a wider range, incorporating both fine-grained measures and register-based measures. The frequency variable tends to be the most linguistically motivated, which enables researchers to explore distinct contributions that different features make to grammatical complexity overall. Therefore, the frequency variable helps research conduct fine-grained investigations on grammatical complexity.

Lu (2011) summarizes 14 measures that performed effectively in L2 writing based on previous influential research studies, i.e., Wolfe-Quintero et al. (1998) and Ortega (2003). I updated Lu's summary based on recent studies in L2 writing where the 14 measures were applied to measure grammatical complexity (see Table 1 below). The updated number of studies is determined by adding the number of studies documented in Lu (2011) and studies collected in the past two decades (1998 - 2018) mostly from the two flagship journals in L2 writing and ESL education, *Journal of Second Language Writing* and *TESOL Quarterly* as well as other popular journals in second language studies, *Journal of English for Academic Purposes*, *English for Specific Purposes*, and *Assessing Writing*. The collected studies are limited to English.

Table 1. The Common Complexity Measures for L2 Writing

| Variables | Measures | Types of Measures | No. of Studies in Lu (2011) | Updated No. in this dissertation |
|---|---|---|---|---|
| Length | Mean length of clause | Length of production | 9 | 12 |
| | Mean length of sentence | Length of production | 10 | 15 |
| | Mean length of T-unit | Length of production | 40 | 44 |
| Ratio | Clauses per sentence | Sentence complexity | 2 | 3 |
| | Clauses per T-unit | Subordination | 18 | 20 |
| | Complex T-units per T-unit | Subordination | 1 | 2 |
| | Dependent clause per clause | Subordination | 3 | 4 |
| | Dependent clause per T-unit | Subordination | 3 | 5 |
| | Coordinate phrases per clause | Coordination | 0 | 3 |
| | Coordinate phrases per T-unit | Coordination | 1 | 3 |
| | T-unit per sentence | Coordination | 5 | 7 |
| | Complex nominals per clause | Particular structures | 0 | 1 |
| | Complex nominals per T-unit | Particular structures | 1 | 3 |
| | Verb phrases per T-unit | Particular structures | 0 | 2 |

The first column in the table presents the two of the four major variables: *length* and *ratio*, because all the 14 measures in Lu (2011) are either length-based or ratio-based; the second column shows the 14 measures summarized in Lu (2011); the third column identifies the types of these measures, i.e., the types of grammatical structures to be measured; the fourth column includes the number of studies that Lu (2011) provided based on Wolfe-Quintero et al. (1998); the fifth column presents the corresponding updated number of studies that applied individual measures in the past two decades (1998 – 2018). As these common measures in Table 1 are only based on length and ratio, the discussion of the two variables below is based on the 14 measures. The other two variables, *index* and *frequency*, are discussed via the measures documented in Wolfe-Quintero et al. (1998) and in studies of L2 writing conducted between 1998 and 2018.

### *2.3.1. Length*

The most frequent measures of grammatical complexity are based upon *length* in L2 writing. However, there is a validity problem: Do length-based measures measure grammatical fluency or grammatical complexity? This problem remains unsolved. Ortega (2003) examined 25 empirical studies regarding grammatical complexity in college-level writing, summarizing that three measures are most frequent: mean length of sentence (MLS), mean length of clause (MLC), and mean length of T-unit (MLTU). Based on Ortega's review, MLTU has been used in 24 out of the 25 studies in research on the topic of college-level L2 writing, which is much more than the other measures of grammatical complexity. MLC has been used in six studies and MLS has been applied in four studies. Table 1 presents the number of studies reported by Lu (2011) and the updated numbers in this dissertation are: MLTU (40, 44), MLC (10, 15), and MLS (9, 12)[4]. Consistent with Ortega (2003), these three measures are much more popular than the others in the general context of L2 writing, except for one measure — clauses per T-unit (C/T) — which was used in 18 and 20 studies, respectively. While many researchers have considered length as a representation of grammatical complexity (e.g., Norris & Ortega, 2009; Ortega, 2003), Wolfe-Quintero et al. (1998) argue that this index measures grammatical fluency[5]. They categorized sentence length, clause length, and T-unit length as fluency measures and provided a number of length-based measures for grammatical fluency, namely, words in T-units, words in clauses, clause length, sentence length, and T-unit length. In contrast, more recent research from the 2000s and 2010s categorizes *length* as a representation of grammatical complexity (e.g., Berlage,

[4] The first numbers in parentheses are the numbers of studies documented in Lu (2011), whereas the second numbers are the updated numbers in this dissertation.
[5] Wolfe-Quintero et al. (1998) describe grammatical fluency as the producing and processing of language in real time, which is related to various aspects, for instance automaticity of language use, length of language output and appropriate of language routines.

2014; Crossley & McNamara, 2014; Lu, 2011; Norris & Ortega, 2009; Ravid & Berman, 2010; Yang et al., 2015). So far, this validity issue has not been clearly addressed by applied linguists and/or scholars in L2 writing.

This validity problem should receive close research attention. Wolfe-Quintero et al. (1998) mention two types of language processing: language access and language representation. Fluency refers to how comfortable an L2 writer automatizes the process of gaining access to language knowledge, whereas complexity refers to how an L2 writer uses knowledge to construct language structures (Wolfe-Quintero et al., 1998). The validity problem of length-based measures is caused by the different purposes these measures are used for. If length-based measures are used to study the automaticity of producing and processing language, then they should be considered fluency measures. If the length-based measures are used to evaluate the representation of language structures, then they should be considered complexity measures. In terms of *length*, what makes the boundary more blurred between grammatical fluency and grammatical complexity is a hypothesis among certain groups of scholars: longer sentences tend to be more complex than shorter sentences because the number of words in sentences are strongly correlated with the number of nodes in their syntactically parsed structures (Berlage, 2014). Overall, *length* may not accurately measure the construct of grammatical complexity, and the length-based measures should be avoided in writing studies and my dissertation as well.

### *2.3.2. Ratio*

Although Norris and Ortega (2009) have already argued that grammatical complexity should be studied as a multi-dimensional construct, most ratio-based measures have still been based on clauses, T-units, and sentences, which primarily measure subordination. The other types of measures (i.e., coordination and phrasal structures) have rarely been applied in L2

writing research. Wolfe-Quintero et al. (1998) mentioned that "[t]here are various types of grammatical complexity ratios that measure the relationship between clauses, sentences, and T-units" (p. 82). They summarize three main types of ratios in particular: (a) the ratio of all clauses to other linguistic units (e.g., clauses per T-unit), (b) the ratio of dependent clauses to other linguistic units (e.g., adverbial clauses per T-unit), and (c) the ratio of coordinate clauses to other linguistic units (e.g., coordinate clauses per T-unit). In Table 1 above, besides the three length-based measures, all the remaining 11 measures are ratio-based, measuring four specific types of grammatical structures:

1. sentence complexity (i.e., clauses per sentence)
2. subordination (e.g., clauses per T-unit)
3. coordination (e.g., coordinated phrases per clause)
4. particular structures in relation to linguistic units (e.g., complex nominals per T-unit).

Taking a close look at the ratio-based measures in Table 1, I noticed two important points. First, only one measure tends to be frequent: clauses per T-unit (18, 20). The rest of the ratio-based measures have not been substantially applied in L2 writing studies. "[T]his is probably due to the lack of adequate computational tools for automatic complexity measurement and the labour-intensiveness of manual computation" (Bulté & Housen, 2012, p. 34). Researchers in applied linguistics have been working on automating the ratio-based measures of grammatical complexity based on computational tools. For instance, Kyle (2016) has recently been developing TAASSC[6], a tool related to grammatical analysis, which can calculate a large number of ratio-based measures[7] (e.g., adverbial clauses per clause). Second, the ratio-based measures on grammatical complexity are subordination-driven. Table 1 shows four types of

[6] TAASSC: https://www.linguisticanalysistools.org/taassc.html
[7] Measure List: https://drive.google.com/file/d/1hHJM- VGp8T037YRrNeiOy0zEW9VyTK1/view

grammatical structures captured by the ratio-based measures, namely, sentence, subordination, coordination, and particular structures. Based on the updated number of studies in the table, subordination measures (31) are dominant, accounting for around 58.49% of the total of ratio-based measures (53) in previous studies. The measures of sentence (3), coordination (13), and particular structures (4) altogether make up 41.50%. The percentages are even more imbalanced based on the number of studies reported in Lu (2011): subordination measures (25) are dominant, accounting for 73.53% of the total of ratio-based measures (34) in previous studies. The measures of sentence (2), coordination (6), and particular structures (1) altogether account for about 26.47%.

As the ratio-based measures are subordination-driven, they support the importance of Norris and Ortega's (2009) argument that grammatical complexity should be studied as a multi-dimensional construct including not only subordination but also coordination and phrasal construction. As an important type of phrasal construction in academic writing, closer attention to NPs is warranted. Although the ratio variable includes two specific measures related to NPs (i.e., complex nominals per clause and complex nominals per T-unit), nominal structures (or nominals) are not the same as NPs. In L2 writing, NPs primarily refer to noun-headed phrases, whereas nominal structures (or nominals) refer to noun-headed phrases, pronoun-headed phrases, and nominal clauses (Biber et al., 2011; Lu, 2011). Therefore, no ratio-based measures in Table 1 are particularly designed for NPs, and future research attention should be paid to the absence of ratio-based measures on NPs.

### *2.3.3. Index*

The index-based *measures* are based on a formula to calculate composite scores comprised of weighted scores of diverse grammatical structures (Wolfe-Quintero et al., 1998).

The difficulty in calculation causes its infrequent use for index-based measures in L2 writing and explains the fact that the effectiveness of index-based measures cannot be easily replicated and triangulated in recent empirical studies. The index-based measures are often composite scores that are argued to represent overall grammatical complexity. For example, Flahive and Snow (1980) used the Complexity Index Score to discriminate grammatical complexity in different stages of L2 writing development. Having reviewed previous studies, Flahive and Snow (1980) built an equation to weigh grammatical structures (see Example 1 below).

**Example 1: The Complexity Index Score**

- *Calculation:* sum of T-unit scores / number of T-units
- *T-unit score*: score of weighted structures plus number of words in T-unit
- *Coding scheme for the scores of weighted structures:*
    - 1 = derivational morphemes & adjectives
    - 2 = relatives, embedded clauses, possessives, comparatives
    - 3 = adverbial and noun clauses

Wolfe-Quintero et al. (1998) summarized three existing index-based measures used in previous research in L2 writing, namely, the Coordination Index in Bardovi-Harlig (1992), the Complexity Formula in Perkins (1980), and the Complexity Index in Flahive and Snow (1980) (see Table 2). The Coordination Index incorporates coordinate constructions, whereas the other two indices measure grammatical complexity in general.

In L2 writing, only a few studies have developed and/or applied index-based measures in the past 20 years. As far as I know, only a certain group of applied linguists with computational background have built a tool (i.e., Coh-Metrix[8]) that can calculate index-based measures to

[8] https://www.memphis.edu/iis/projects/coh-metrix.php

explore grammatical complexity in L2 writing. For instance, Crossley and McNamara (2014) applied 11 Coh-Metrix indices of clausal and phrasal complexity to analyze the grammatical development. These indices were all automatically computed, which include indices of syntactic variety, syntactic embeddings, and incidence of phrase types.

Table 2. Indices for Grammatical Complexity in L2 Writing Research

| | Measures | Construct | Publication | No of Studies |
|---|---|---|---|---|
| Index | Coordination Index | Coordination | Bardovi-Harlig (1992) | 3 |
| | Complexity Formula | Complexity in general | Perkins (1980) | 1 |
| | Complexity Index | Complexity in general | Flahive and Snow (1980) | 8 |

*Note*. Adapted from Wolfe-Quintero et al. (1998).

Although the three index-based measures in Table 2 were designed based on research findings in language studies, "none of the three grammatical complexity indices proposed by researchers significantly related to second language development" (Wolfe-Quintero et al., 1998, p. 94). Two reasons may cause the ineffectiveness of the index-based measures. First, due to the complex process of generating a composite index score, the measures have not been replicated in substantial studies. The relation between the measures and writing development cannot be thoroughly explored, and previous findings cannot be effectively triangulated. Second, it is reasonable to assume that computational techniques in or before the 1990s were not advanced enough to help researchers apply the index-based measures to a large number of texts. There might be biases of writing samples in previous studies, and the biases might cause the ineffectiveness of the index-based measures. Overall, little empirical evidence so far can support the effectiveness of using index-based measures to investigate writing development. Therefore, scholars should be careful to apply them in writing research.

### *2.3.4. Frequency*

Measures based on the frequency variable are often a set of grammatical features. Wolfe-Quintero et al. (1998) summarized 15 grammatical features used in previous research about L2 writing development, which included: four lexical features (i.e., adjectives, pronouns, articles, and connectors), one phrasal feature (i.e., prepositional phrases), five clausal features (i.e., reduced clauses, dependent clauses, adjectival clauses, adverbial clauses, and nominal clauses) and five other features (e.g., transitional connectors). Since the 2010s, researchers have begun integrating a large set of grammatical features with the support of computational techniques (e.g., tagging and parsing). For instance, Grant and Ginther (2000) studied 20 features; Biber et al. (2011) explored 28 features; Staples et al. (2016) investigated 23 features. In all these empirical studies, the frequencies of grammatical features were calculated individually, so researchers were able to analyze the contribution of different grammatical features to grammatical complexity.

The application of *frequency* has experienced a change in writing research. Raw frequencies of grammatical features were substantially used in L2 writing research until the 1990s, but the validity of using raw frequencies is doubtful because of "the lack of a fixed delimiter" (Wolfe-Quintero et al., 1998, p. 75). To be more specific, measures based on raw frequencies are often influenced by text length. For instance, a 1,500-word essay is very likely to have more nouns than a 950-word essay. In response to this situation, normed frequency has been increasingly applied in L2 writing studies in order to avoid the influence of text. Normed frequency means the frequency of a grammatical feature which is standardized to a certain text length (e.g., 1,000 words, 10,000 words).

**Example 2**

- *Calculation:* normed frequency = (raw frequency/length) * normed length

Applied/corpus linguists have used normed frequency in many L2 writing studies. Example 2 illustrates the formula of normalizing raw frequencies. For instance, if the raw frequency of relative clauses were 6 in an 876-word paper and we set the normed length as 1,000 words, then the normed frequency would be: (6/876) *1,000 = 6.85. The normed length can be set as any number, but in most cases, it should be a number close to mean length of the files in a corpus.

In addition, the use of frequency-based measures have also changed during the process of selecting different grammatical features to represent grammatical complexity. In L2 writing, there are two types of frequency-based measures: (1) fine-grained measures and (2) register-based measures. Although both types of measures are applied by drawing on a number of grammatical features, the selection of the features are different. Researchers in register analysis select grammatical features for their research based on the situation of language use (i.e., register), whereas this is not the case for researchers who applied fine-grained measures. From the perspective of register analysis, grammatical features have different functions in different registers (Biber & Conrad, 2009). Considering the influence of registers, Biber et al. (2011) propose a developmental index of writing complexity features in writing, which includes different subforms for particular grammatical features. For instance, *prepositional phrase* includes prepositional phrases headed by *of*, prepositional phrases headed by other prepositions, and prepositional phrases followed by -ing clauses. This is mainly due to the fact that different types of prepositional phrases tend to play different roles in writing development. Recently, an increasing number of studies have applied register-based measures in L2 writing to explore grammatical complexity (e.g., Biber et al., 2011; Lan, Lucas, & Sun, 2019; Parkinson & Musgrave, 2014; Staples et al., 2016; Wang & Beckett, 2017).

The growing use of register-based measures in writing research is ascribed to the core value that frequency-based measures added to the study grammatical complexity, i.e., the fine-grained investigation of individual grammatical features. Researchers in applied linguistics have argued that the composite nature of the measures is based on the three aforementioned variables: *length*, *ratio*, and *index*. Biber, Gray, & Staples (2016) point out that there have been "a few holistic measures designed to capture the entire system of grammatical complexity", but these composite measures "confound the analysis of multiple grammatical features that have distinct functions and distributions" (p. 649). This indicates that the uniqueness of grammatical features is not well represented via *length*, *ratio*, and *index*. For instance, clauses per T-unit (a ratio-based measure) cannot distinguish different types of clauses (e.g., adjectival clauses, nominal clauses, and adverbial clauses). In contrast, frequency-based measures are linguistically motivated. The inclusion of individual grammatical features allows for a fine-grained investigation to demonstrate contributions that different grammatical features make to grammatical complexity. Between the two types of frequency-based measures (i.e., fine-grained measures and register-based measures), register-based measures are proposed based on functions of grammatical forms (and subforms) in certain registers (e.g., academic writing). Compared to fine-grained measures, register-based measures provide more information for researchers to analyze grammatical complexity in writing research, which also helps researchers to better interpret their research findings.

## 2.4. Grammatical Complexity – an Evolving Concept

Overall, this chapter has reviewed the changes in how grammatical complexity has been discussed and understood over the past 20 years in writing research (primarily L2 writing). The review demonstrates that this construct has experienced changes along the conceptual and the

operational dimensions. In terms of conceptualization, although there is still no consensus on the definition of grammatical complexity, two perspectives co-exist in writing research: grammatical complexity has been theorized as (a) an independent construct with systemic and structural complexities and (b) as a dependent construct based on registers. Also, the representation of grammatical complexity has been expanded from clausal complexity to clausal and phrasal complexity. As a critical and frequent type of phrasal feature, NPs need to be integrated into the representation of grammatical complexity. According to the operationalization, the measures of grammatical complexity are based on four major variables (i.e., length, ratio, index, and frequency), which have experienced corresponding changes. Generally, the measures based on *length*, *ratio*, and *index* tend to be composite, whereas the measures based on *frequency* should be applied if researchers intend to explore how each unique grammatical feature contributes to grammatical complexity. For NPs, register-based measures (i.e., a type of frequency-based measure) help researchers conduct fine-grained examinations of individual grammatical features that contribute to NP complexity, and research findings can also be better interpreted because these grammatical features are proposed based on specific registers (e.g., academic writing).

# CHAPTER 3. NOUN PHRASE COMPLEXITY

This chapter firstly presents a selection of NP definitions (i.e., noun-headed phrases) based on the existing scopes of NPs in applied linguistics. NPs can refer to different grammatical structures in different fields (e.g., L2 writing, syntax), so it is important to clarify this term in this dissertation. Then, NP complexity is discussed from the three perspectives of grammatical complexity discussed in Chapter 2 (i.e., the systemic perspective, the structural perspective, and the register perspective). I select the register perspective to study NP complexity after reviewing corresponding quantifiable measures for their strengths and weaknesses. Last, a linguistic model of NP complexity is built based on Biber et al.'s (2011) hypothesized developmental index of writing complexity features. This linguistic model is presented at the end of this chapter and is used in the remaining chapters of the dissertation.

## 3.1. Selection of NP Definition

NPs can refer to multiple grammatical structures in linguistics (e.g., nouns, pronouns, and nominal structures). In this dissertation, I define NPs only as noun-headed phrases and exclude other possible grammatical structures. In the *Longman Grammar of Spoken and Written English*, Biber, Johansson, Leech, Conrad and Finegan (1999) identify two scopes for NPs: a strict scope and a broad scope. According to the strict scope, an NP refers to a head noun, either alone or with determiners and modifiers. For the broad scope, an NP frequently refers to "any unit which appears in the positions characteristic of noun-headed structures, including clauses" (p. 97). Therefore, this term (i.e., "NP") can include three types of constructions: noun-headed phrases,

pronoun-headed phrases, and nominal clauses[9] (Biber et al., 1999; Hunston & Francis, 2000). Example 3-A presents an NP with the head noun, *method*, accompanied by a determiner (*a*), a premodifying noun (*teaching*) and a prepositional phrase as a post-noun modifier (*in bilingual literacy*). Example 3-B displays an NP with the head pronoun (*those*) and a relative clause as a post-noun modifier (*who are interested in corpus linguistics*). Example 3-C presents an NP with a wh-clause as the head noun (*what he argued in the speech*) in the subject position.

**Example 3**

A. a teaching [method] in bilingual literacy

B. [those] who are interested in corpus linguistics

C. [what he argued in the speech] does not solve my problem

I apply the strict linguistic scope of NP (i.e., noun-headed phrases) rather than all three types of construction shown in Example 3. Two reasons support this linguistic choice. First, register studies in corpus linguistics show that pronouns are rarely accompanied by modifiers in academic writing (Biber et al., 1999). In English, there are only a few exceptions, for instance *those* (e.g., *those who violate cultural norms*) and *one* (e.g., *a special one that includes two models*). Second, applied linguists have noted that nominal clauses in academic writing are often reduced to noun-headed phrases (Wolfe-Quintero et al., 1998). For instance, the sentence in Example 3-C can be reduced to "*His argument in the speech does not solve my problem"*. This is because complex noun phrases (i.e., noun-headed phrases with pre- and/or post-noun modifiers) can pack intensive information in a compact space (Biber et al., 2011), which is a characteristic of academic writing. Thus, pronoun-headed phrases and nominal clauses are excluded, and the term (NP) only refers to noun-headed phrases in this dissertation.

[9] Nominal clauses refer to constituents that are used in the position of a subject, subjective predicative, or direct object.

Biber et al. (1999) propose a universal and canonical structure for NPs (i.e., noun-headed phrases): *determiner + premodifier(s) + head noun + postmodifier(s)* (see Example 3-A). A *head noun* is obligatory in an NP, but a *determiner*, *premodifier(s)*, and/or *postmodifier(s)*[10] are optional. As grammatical complexity includes three specific perspectives (i.e., the systemic perspective, the structural perspective, and the register perspective), I will then use these perspectives to summarize how NP complexity is operationalized to quantifiable measures in writing studies based on the canonical NP structure. This summary helps me to finalize the most appropriate perspective of NP complexity that I should follow in this dissertation.

### 3.2. Selection of NP complexity

In the following sections, I provide detailed evaluations of systemic NP complexity, structural NP complexly, and register-based NP complexity. The operationalized measures of these three types of NP complexity are presented, and their strengths and weaknesses are discussed with specific examples. Based on careful evaluations, I select register-based NP complexity for this dissertation, which is operationalized as the 11 noun modifiers identified in Biber et al.'s (2011) hypothesized developmental index of writing complexity features. Compared to measures based on systemic NP complexity and structural NP complexity, the measures (i.e., the 11 noun modifiers) of register-based NP complexity have three major strengths: (1) they were developed based on academic writing, the same register that this dissertation investigates; (2) they were presented based on the developmental stages of academic writing in Biber et al.'s index, which provides important information for NP analysis in academic

[10] A *postmodifier* in the traditional sense "implies the functional distinction between modifiers and complements", but the term is also often used to denote "all structural elements following the head of the NP" in recent literature (Berlage, 2014, p.1). Therefore, *postmodifier* covers the grammatical structures functioning as both modifiers and complements following head nouns in this dissertation.

writing; (3) they have not been criticized for their validity, reliability, or foundation of operationalization in existing literature.

### *3.2.1. Systemic NP Complexity*

Systemic NP complexity refers to the breadth of NP patterns. There are two types of measures for the breadth of NP patterns in previous literature. The first type of measure is based on the classification of NP patterns via specific coding schemes, while the second type of measure maintains fine-grained distinctions among noun modifiers. A few studies in writing research have been conducted with the first type of measure. The systems for classification of NP patterns are difficult to manage because various NP patterns cannot easily be categorized into the coding schemes. In contrast, the second type of measure (i.e., fine-grained measures) has been widely applied in writing research because fine-grained measures allow easier identification of noun modifiers. These noun modifiers are sufficiently linguistically motivated, which has implications for instruction in writing classes.

In terms of the first type of measure, Ravid and Berman (2010) conducted a study to explore the development of NP complexity by students at different academic levels from primary school to graduate school. They included a specific criterion named syntactic variability of NPs, which is related to the number of different types of noun-modifying categories. Ravid and Berman (2010) provide specific examples for their coding scheme[11]:

- "The NP [*a conflict with me*] has only one type of modifier, the prepositional phrase *with me*, and so rates the lowest score of 1.
- The NP [*one belief that opposes the belief of another nation*] was assigned a score of 2 because it contains two different types of modifiers: the quantifier *one* and the relative

[11] The format was adjusted to make the visual presentation of the coding scheme clear.

clause *that opposes the belief of another nation* (which in itself contains two NPs *the belief* and *another nation* in a possessive construction).

- Also, assigned a score of 2 on this criterion was the NP [*very low tolerance for other people*] because it has an adjective *low* and a prepositional phrase *for other people*.
- The NP [*every realm from the most internal and personal to the most public and universal*] rated a score of 3 since it contains the correlative coordinating construction of prepositional phrases *to . . . from* with two coordinated adjectives *most internal and personal and public and universal in each*.
- The (also very rare) score of 4 was given to variegated NP modification as in [*many different kinds of extremely valuable contributions to the wellbeing of people living and working in Tel Aviv as the city of their choice*]" (p.11).

So far, this type of ranking has rarely been applied to NP complexity in L2 writing, because it is extremely hard to categorize all possible cases of NPs based on the coding scheme as shown above. For instance, should a NP as in "*a useful and common research method*" receive a score of 1 or 3? This NP contains three premodifiers with a coordinated structure: two attributive adjectives (*useful, common*) and a premodifying noun (*research*). It could receive a score 3 because it has three modifiers, but this NP is intuitively comparable in complexity to the example with score 1 above. Therefore, the first type of measure used in Ravid and Berman (2010) can be problematic to apply to large-scale corpus-based studies that include various NP patterns.

The second type of measure is based on fine-grained grammatical features related to NP complexity (i.e., nouns, premodifiers and postmodifiers). Studies with fine-grained measures often include multiple types of premodifiers and postmodifiers, and researchers then calculate their frequencies to represent NP complexity. For example, Parkinson and Musgrave (2014) included a set of grammatical features to investigate NP complexity in academic writing (e.g., attributive adjectives, relative clauses, prepositional phrases as postmodifiers). In contrast to the first type of measure (i.e., the classification of NP patterns), fine-grained measures have been utilized more frequently in writing studies. Scholars prefer fine-grained measures because only

certain set of grammatical features can modify nouns in English. A convenient way of using fine-grained measures is to include nouns and all the possible noun modifiers in a study. Also, fine-grained measures can reveal more detailed linguistic information, presenting the uniqueness of individual noun modifiers. Thus, pedagogical implications can be derived from these measures, which is an important research purpose for studies in applied linguistics.

### *3.2.2. Structural NP Complexity*

Structural NP complexity refers to the depth of NP construction. Little research has attempted to give an exploration of the depth of NP construction (Berlage, 2014). In theoretical linguistics, the depth of NP construction is measured based on two variables: *length* and *structure* (Berlage, 2014). *Length* is operationalized as the number of words in an NP, and *structure* is operationalized as node counts (or phrasal node counts) in a syntactically parsed NP. However, a major problem is that the results of the length-based measures and node-based measures do not always align. This reveals a potential problem between length-based measures and node-based measures. Berlage (2014), who conducted an empirical study on NP complexity in theoretical linguistics, is considered a comprehensive publication on NPs in English. Berlage included two variables in the study to explore NP depth: *length* and *structure*. According to the use of these two variables, Berlage (2014) proposed a hypothesis:

- Longer NPs are syntactically more complex than shorter NPs.

The relationship between *length* and *structure* has been challenged by Miller and Chomsky (1963), who argued that a shorter sentence could be syntactically more complex than a longer sentence. Miller and Chomsky provided concrete evidence for their argument based on their *Metric of Structural Complexity*. The metric score proposed in Miller and Chomsky (1963) is computed as the number of non-terminal nodes divided by the number of terminal nodes. The

metric scores suggest the degree of syntactic complexity: the larger the metric score, the more structurally complex a sentence is. Example 4 below illustrates the computing process in the setting of measuring NP complexity. Example 4-A shows an NP with six words, *all the leaders of the riot*, and Example 4-B presents an NP with four words, *the man I saw*.

**Example 4**

A. They brought [all the leaders of the riot] in

B. They brought [the man I saw] in

Note. The examples were used in Miller and Chomsky (1963)

Syntactic parsing of Example 4-A

```
(ROOT
  (NP1
    (NP2 (PDT all) (DT the) (NNS leaders))
    (PP (IN of)
      (NP3 (DT the) (NN riot))))
```

Token = 6

Non-terminal nodes: ROOT, NP1, NP2, PP, NP3

Terminal nodes: PDT, DT, NNS, IN, DT, NN

Syntactic parsing of Example 4-B

```
(ROOT
(NP1 (DT the) (NN man)
  (SBAR
    (S
      (NP2 (PRP I))
      (VP (VBD saw)))))))
```

Token = 4

Non-terminal nodes: ROOT, NP1, SBAR, S, NP2, VP

Terminal nodes: DT, NN, PRP, VBD

In Example 4-A, the parsing result above shows that there are five non-terminal nodes and six terminal nodes, so the metric score is 5/6 = 0.83. In Example 4-B, the parsing result demonstrates that there are six non-terminal nodes and four terminal nodes, so the metric score is 6/4 = 1.50. In this case, the short NP in 6-B (token = four words) is in fact structurally more complex than the long NP in 6-A (token = six words). The hypothesis that longer sentences are more complex has been challenged by Miller and Chomsky's metric score. Longer NPs are not always syntactically more complex than shorter NPs. Therefore, the inconsistency between *length* and *structure* indicates that there might be a problem between length-based measures and node-based measures.

Other problems of length-based measures and/or node-based measures are also summarized in the list below: Both types of measures are composite in nature. NP complexity is only represented by the numbers of words or the numbers of nodes. These measures do not allow researchers to have a fine-grained investigation on how different grammatical features contribute

to NP complexity in writing. As Chapter 2 mentioned, the length variable has a validity problem. There is still an ongoing debate on whether *length* is to measure grammatical fluency or grammatical complexity. There might be a validity issue when to use *length* to measure NP complexity. In practical context, node-based measures are labor-intensive to be applied. Syntactic parsers are built in some computational tools, such as Stanford CoreNLP[12] and Python Natural Language Toolkit[13], but, to the best of my knowledge, no program has been developed for node counting. It is labor-intensive to count all nodes manually in writing studies. Overall, we should be careful when these two types of measures are applied in writing research.

### *3.2.3. Register-based NP complexity*

In contrast to systemic and structural NP complexity, register-based NP complexity considers the influences of register on NPs. The measures of register-based NP complexity are specific grammatical features (i.e., noun modifiers). Register-based NP complexity is operationalized as the 11 noun modifiers identified in Biber et al.'s (2011) developmental index of writing complexity features. The noun modifiers are hypothesized in relation to academic writing, consistent with the register in this dissertation. The noun modifiers are also categorized with respect to hypothesized stages of writing development.

In academic writing, corpus linguists have revealed that approximately 60% of NPs are complex NPs, with either premodifiers and/or postmodifiers (Biber et al., 1999). The use of NPs in academic writing supports the importance of analyzing NP complexity in this specific register. Biber et al. (2011) examined a large set of grammatical features in conversations and research articles. They then proposed a developmental index of writing complexity features, which

[12] https://stanfordnlp.github.io/CoreNLP/
[13] https://www.nltk.org/

includes 11 specific noun modifiers (see Table 3). The noun modifiers were then associated with different stages of writing development from Stage 2 to Stage 5. The 11 noun modifiers were also categorized into different types based on their positions and grammatical types. According to the positions, there are two premodifiers (i.e., attributive adjectives and premodifying nouns), and the remaining nine are postmodifiers.

Table 3. Noun Modifiers in Biber et al.'s (2011) Index

| Stage | Noun Modifier | Position | Type | Example |
|---|---|---|---|---|
| 1 | N/A | | | |
| 2 | attributive adjectives | pre | phrasal | an effective solution |
| 3 | premodifying nouns | pre | phrasal | a research method |
| | relative clauses | post | clausal | the book which she gave me |
| 4 | -ing clauses | post | clausal | the man holding the flag |
| | -ed clauses | post | clausal | the notes taken by students |
| | prepositional phrases (*of*) | post | phrasal | the policy of immigration |
| | prepositional phrases (other) | post | phrasal | the laptop on the table |
| 5 | prepositions + ing clauses | post | phrasal | the ability of using computers |
| | noun complement clauses | post | clausal | the fact that the price was dropped |
| | infinitive clauses | post | clausal | the way to debug the program |
| | appositive noun phrases | post | phrasal | political corruption, a big problem |

*Note*. this table was first built by Lan and Sun (2019) based on Biber et al. (2011). Stage 1 is the beginning stage of writing development, and Stage 5 is the advanced stage of writing development.

For grammatical types, six of them are phrasal, namely, attributive adjectives, premodifying nouns, prepositional phrases (of), prepositional phrases (other), and appositive noun phrases. Five are clausal, including relative clauses, -ing clauses, -ed clauses, noun complement clauses, and infinitive clauses. Considering the grammatical functions of the noun modifiers in academic writing, Biber et al. (2011) included sub-forms of one specific type of noun modifier: prepositional phrases as postmodifiers. These were further categorized as prepositional phrases headed by *of*, prepositional phrases headed by other prepositions, and prepositional phrases followed by an -ing clause. This move was very important for writing

research because prepositional phrases (*of*) have many more grammatical functions than prepositional phrases (*other*), such as to construct partitive structures (e.g., *part of*) and to build quantitative structures (e.g., *a number of*). In addition, prepositional phrases followed by -ing clauses is a noun modifier that tends to be used more in advanced stages of writing development. Prepositional phrases followed by -ing clauses should also be distinguished from regular prepositional phrases as noun modifiers. In L2 writing and academic writing, the 11 noun modifiers have been applied in some recent studies to explore NP complexity in L2 writing (e.g., Lan, Lucas & Sun, 2019; Staples et al., 2016; Parkinson & Musgrave, 2014; Wang & Beckett, 2017).

In conclusion, I consider register-based NP complexity the most appropriate for my dissertation, and three reasons support this choice. First, the measures of register-based NP complexity include specific grammatical features that reveal their unique contributions to NP complexity, which is different from composite measures (i.e., length-based and node-based measures). Second, the measures of register-based NP complexity are frequencies of specific grammatical features (e.g., nouns and noun modifiers). Compared to coding schemes with regard to the classification of NP patterns, the frequencies of the noun modifiers are less time-consuming to calculate and can be automated with a high precision rate. Third, in contrast to fine-grained measures, register-based measures are selected based on certain registers (e.g., academic writing), which provides more information for interpreting the findings of this dissertation.

# CHAPTER 4. NOUN PHRASES IN ACADEMIC WRITING

This chapter explores the roles of NPs in academic writing. The first section explains the importance of NPs in academic writing. The second section presents register studies related to NPs in academic writing. The third section discusses research on the relationships among the use of NPs, writing proficiency, and writing development. Last, research gaps are summarized, and research questions are listed.

## 4.1. The Importance of NPs in Academic Writing

The importance of NPs can be interpreted from three angles. Complex NPs (i.e., NPs with multiple modifiers) are frequent in academic writing. It is often the case that nouns carry the key information of sentences in the subject and object positions, and modifiers in complex noun phrases perform the functions of modification and complementation, effectively expanding the meaning of head nouns (Biber et al., 2011). Via corpus-based methods, Biber et al. (1999) demonstrated that nearly 60% of NPs in academic writing are complex NPs, with pre- and/or post-modifiers. Also, in academic contexts, writing frequently uses nouns with phrasal modifiers, which is referred to as "compressed NPs". The dense use of compressed NPs generates a specific style (i.e., nominal style) in academic writing. Although some have argued that the nominal style may interfere with clarity, this style contrasts with the verbal style of conversation and makes academic writing specialized, professional, and technical (Fang, Schleppegrell, & Cox, 2006; Sebeok, 1960). Corpus-based studies demonstrate that compressed NPs have been increasingly used in academic writing from the 17$^{th}$ to the 20$^{th}$ centuries (Biber & Clark, 2002; Biber & Gray, 2011; Biber & Gray, 2013), which may suggest a preference for the nominal style in the academic community.

Last, phrasal features (NPs) have not been widely selected to represent grammatical complexity until the recent years (see Chapter 2) in L2 writing, which overshadowed researchers' understandings of writing development. The development of grammatical competence is a nonlinear process, symbolized as the shape of omega (Ω) (Wolfe-Quintero et al., 1998; Norris & Ortega, 2009). Scholars in applied linguistics have categorized this development in three general stages:

1. using simple and coordinate structures to construct sentences
2. embedding subordinate (compound and complex) structures in sentences
3. producing compressed noun phrases with coordinate and phrasal features to pack intensive information (Biber et al., 2011; Parkinson & Musgrave, 2014; Staples et al., 2016; Wolfre-Quintero et al., 1998).

Without NPs, researchers can only capture the writing development from stage 1 to stage 2, which is based on the increasing use of clausal features (e.g., subordinate clauses), and writing development from stage 2 to stage 3 cannot be effectively explored in the analysis of grammatical complexity. Thus, the importance of NPs in academic writing comes from the high frequency of NPs, the nominal writing style in academic contexts, and the limitation of understanding writing development without integrating NPs into representations of grammatical complexity.

### 4.2. Register Studies Related to NPs

A number of register studies have been conducted using corpus-based or corpus-driven analyses of NPs in written registers, including academic writing (e.g., Biber, 1988; Biber, 1992; Biber & Clark, 2002; Biber & Gray, 2011, Biber & Gray, 2013). These register studies can be

categorized into two groups: (a) investigating the historical trends of NPs use in academic writing and (b) comparing the use of NPs in written and spoken language.

### *4.2.1. The Trend of Using Compressed NPs in Academic Writing*

Applied linguists have researched the changes in the use of pre-noun modifiers and post-noun modifiers in written registers across different periods of time (e.g., Biber & Clark, 2002; Biber & Gray, 2011). One important finding is that phrasal modification in NPs has become much more frequent in the 20th century than in previous centuries (i.e., the 18th and 19th centuries). Increased use of phrasal noun modifiers is due to the fact that phrasal modifiers (e.g., premodifying nouns, prepositional phrases, and appositive NPs) generate compressed NP constructions, which can carry intensive information in a compact space. Example 5 includes an appositive NP (*Acrobasus Vaccinii Riley*), an attributive adjective (*major*), and a prepositional phrase (*of Vaccinium spp*).

**Example 5**

- The cranberry fruitworm, Acrobasus Vaccinii Riley, is a major pest of Vaccinium spp in the eastern U.S.A. (in Biber & Conrad, 2009).

Biber and Clark (2002) conducted a corpus-based study, showing the shifts of NP structures based on modifying patterns from the 17th century to the 20th century across four different registers. The corpus (0.7 million tokens) in the study was a subset of a large corpus, Representative Corpus of Historical English Registers (ARCHER), representing two popular registers (i.e., drama and fiction) and two written registers (i.e., news report and medical research papers). Seven noun modifiers were investigated for their frequency changes, including the four phrasal modifiers (i.e., attributive adjectives, premodifying nouns, prepositional phrases (*of*), and prepositional phrases (*other*) and the frequencies of the three clausal modifiers (i.e., -ed clauses, -

ing clauses, and relative clauses). Biber and Clark (2002) found that there was a sharp increase in the use of all four phrasal modifiers over the past century (i.e., the 20th century) in medical research papers, whereas clausal modifiers remained relatively stable across the four centuries. A qualitative analysis revealed that in the medical research papers (1) premodifying nouns were as frequent as attributive adjectives, which was considered the most frequent premodifiers, and (2) in terms of postmodifiers, prepositional phrases were much more frequent than the clausal modifiers (e.g., relative clauses and -ing clauses). Thus, the study illustrated the growing grammatical use for phrasal modification in the academic written register in the past three centuries.

Biber and Gray (2011) continued the investigation of NPs in research articles, a specific academic written register. In addition to the four phrasal modifiers in Biber and Clark (2002), two additional features were also included (i.e., nominalizations and appositive NPs). The six phrasal modifiers were studied in a corpus (1.7 million tokens) of research articles across the 18th, 19th, and 20th centuries. Consistent with Biber and Clark (2002), Biber and Gray (2013) found that the frequencies of all six phrasal modifiers had sharp increases in the 20th century. A functional interpretation showed an expansion of grammatical functions for premodifying nouns and prepositional phrases. For premodifying nouns, richer semantic relationships were observed between the head nouns and the premodifiers than the three common semantic types, i.e., title nouns (e.g., *Doctor Thompson*), location nouns (e.g., *city hall*), and concrete nouns (e.g., *iron chain*). These additional semantic types identified in the study included, for example, institutions/person (e.g., *government officer*), topic identification (e.g., *psychology lecture*), and regulation/administration (e.g., *research fund*). Regarding prepositional phrases, a shift was found from modifying head nouns with only concrete/locative meanings (e.g., *kids from*

*California*) to modifying head nouns with both concrete/locative and abstract meanings (e.g., *improvement in research method*). Thus, this study also shows an increased frequency of phrasal modification on NPs and interprets the grammatical functions of phrasal modifiers in research articles.

Biber and Gray (2013) also investigated how compressed NPs with phrasal modifiers might interact with verb phrases (VPs) in academic writing with the ARCHER corpus. The corpus included three academic written registers, namely, science research articles (1.8 million tokens), non-science research articles (1.9 million tokens), and popular science articles (0.28 million tokens). The features related to the compressed NPs included nouns, nominalizations, and premodifying nouns, whereas the features related to VPs included verbs and verb phrases in different tenses and aspects. Based on a descriptive analysis, Biber and Gray demonstrated that nouns and premodifying nouns have been the most frequent, with the sharpest increase in science research articles over the three centuries. In contrast, simple verbs and verb phrases (with modals, perfect aspect, and verbs) have been the least frequent, with the sharpest decrease in science research articles. A qualitative analysis showed that the pattern of "copular verb + adjective" had been expressed by verbless NPs with phrasal modifiers, namely, "adjectives + nouns" and "nouns, appositive NPs". The study provided additional empirical evidence for the grammatical preference of NPs with phrasal modification in academic writing.

### *4.2.2. Comparison of Noun Phrases between Spoken and Written Registers*

Another group of register studies have supported the proposition that the use of compressed NPs is unique to written registers via *factor analysis*, which they refer as *multi-dimensional analysis*, on a set of lexico-grammatical features. Biber (1986) pointed out that "earlier studies, using a more restricted analysis on few text types and few linguistic features,

reached contradictory global conclusions concerning the relations among spoken/written text types" (p. 410). Biber (1988) conducted an exploratory factor analysis (EFA) to examine the use of 67 lexico-grammatical features in a large corpus (1 million tokens). The corpus covered two subcorpora to represent the spoken register (e.g., face-to-face conversation and broadcasts) and the written register (e.g., press reportage and academic prose). A five-factor model was built in Biber (1988) based on 33 out of 67 features to illustrate the factors that influence linguistic variation across the two registers. He found that certain lexico-grammatical features co-occur in each dimension. In term of NPs, Dimension 1 (Involved vs. Informational production) revealed a co-occurring pattern among nouns, attributive adjectives, and prepositions in informational written register. However, the spoken registers were found to have a co-occurring pattern among completely different grammatical features, including pronouns, wh-clauses, and causative subordinations. Based on this new statistical approach, the study provided initial empirical evidence that compressed NPs are unique to written registers.

Biber (1992) conducted a confirmatory factor analysis (CFA) to examine the goodness-of-fit of the five-factor model reported in Biber (1988). The model was applied to a large corpus (about 1 million words) that included written registers (e.g., academic writing and professional letters) and spoken registers (e.g., conversations and speeches). The confirmatory analysis presented that the goodness-of-fit was good: the GFI[14] was 0.73, meaning that the fit was 73% between a hypothesized model (i.e., the aforementioned five-dimension model) and the actual model (i.e., the factor model in this study). Similar to the finding in Biber (1988), this study revealed co-occurring patterns in written registers (e.g., academic writing). The co-occurring features included nouns, attributive adjectives, and prepositions, with all their factor loadings[15]

[14] Good-of-fit Index (GFI) evaluates the goodness-of-fit of the models built based on factor analysis.
[15] A factor loading greater than 0.30 is considered a significant loading (Brown, 2015).

greater than 0.50. This result provided additional evidence that compressed NPs are an important and unique grammatical feature of written registers.

As the five-factor model was confirmed to have a good goodness-of-fit in Biber (1992), Biber, Conrad, Reppen, Byrd, and Helt (2002) then applied the model to the university setting. The corpus used in the study was the TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL), which covered an academic spoken corpus (about 1.7 million tokens) and an academic written corpus (about 1 million tokens). The result was consistent with previous work. The factor of Involved vs. Information Production demonstrated that nouns, attributive adjectives, and prepositions were used together in university written registers, such as academic writing, textbooks, and course packs. These university written registers "are extremely informational in purpose and are produced under highly controlled and edited circumstances" (Biber et al., 2002, p. 26). Thus, the study found that the dense amount of information packed in compressed NPs is characteristic of these written registers, and it also revealed the language (English) that L2 students would encounter when they entered universities (Biber et al., 2002).

Biber, Gray and Staples (2016) conducted a research to explore how previous factor model of register variation can be used to predict spoken and written task-type variation in TOEFL iBT. Their corpus consists of 2,378 texts to represent four specific task types: spoken independent tasks, spoken integrated tasks, written independent tasks, and written integrated tasks. The five-factor modal in Biber (1988) were then applied to investigate co-occurrence patterns of lexico-grammatical features in their corpus. The factor of Literate vs. Oral Response showed that nouns, premodifying nouns, attributive adjectives, prepositional phrases, noun + of phrases, and passive structures co-occur in written tasks, whereas verbs, verb complement clauses, finite adverbial clauses co-occur in spoken tasks. Similar to Biber et al. (2002), Biber,

Gray and Staples also found that the lexico-grammatical features related to structural compression (i.e., compressed NPs) tend to be used together in written registers in the context of TOEFL iBT.

Register analysis has also been conducted based on other statistical methods to explore lexico-grammatical features in spoken and written registers. Biber, Gray, and Poonpon (2011) examined the use of 28 lexico-grammatical features in a corpus of academic writing (about 2.9 million tokens) and a corpus of conversation (about 4.2 million tokens). The academic writing corpus included journal research articles in biology, education, history, medicine, and psychology. The conversation corpus was compiled from the face-to-face conversation transcripts in the Longman of Spoken and Written Corpus (LSWC). An ANOVA analysis was applied to test the significance of the differences on the 28 features between the two registers. The results demonstrated that there was a significant difference for non-clausal features between academic writing and conversation, including attributive adjectives, premodifying nouns, and prepositional phrases as postmodifiers. Biber et al. (2011) proposed that writers tend to produce more phrases, mostly NPs with phrasal modifiers, in the advanced stage of academic writing.

The studies that have compared written and spoken registers added empirical evidence to support the proposition that compressed NPs are an important and unique grammatical feature of academic writing, which has not been observed in spoken registers. Biber and Gray (2010) summarized that the features related to structural elaboration (e.g., complement and adverbial clauses) are much more common in conversation, whereas the features related to structural compression (e.g., attributive adjectives, premodifying nouns, and prepositional phrases as postmodifiers) are much more common in academic writing (i.e., journal articles). This

challenges the stereotype that writing is structurally elaborate with embedded clauses to provide detailed information. In fact, academic writing relies on the use of compressed NPs.

### 4.3. Writing Development, Writing Proficiency, and NPs

Writing development in academic contexts has been a popular topic in applied linguistics since the 1990s. To better understand writing development from the linguistic perspective, applied linguists began to build developmental indices with a set of grammatical features that might be applied as a yardstick to measure development in L2 writing (Larsen-Freeman, 1983). The contribution of developmental indices is to offer (1) an objective assessment to measure writing development, (2) a precise description of the writing developmental levels which language learners belong to, and (3) potential guidelines for grammatical instruction in the writing classroom (Larsen-Freeman, 1983; Wolfe-Quintero et al., 1998). The measures of writing development have been investigated in both cross-sectional studies (e.g., Biber, Gray, & Poonpon, 2011; Lu, 2011; Parkinson & Musgrave, 2014; Staples et al., 2016; Wang & Beckett, 2017) and longitudinal studies (e.g., Crossley & McNamara, 2014; Mazgutova & Kormos, 2015; Ravid & Berman, 2010; Spoelman & Verspoor, 2010).

However, when the measures of writing development are applied to evaluate writing proficiency, there is a validity issue because it is problematic to assume that "developmental measures and proficiency measures are tapping into the same construct" (Wolfe-Quintero et al., 1998, p.8). Even though both writing development and proficiency are associated with language use across similar linguistic levels, such as vocabulary, syntax, semantics, discourse, and pragmatics (Lu, 2011), so far it is still "not clear whether developmental level can be equated with proficiency level" (Wolfe-Quintero et al., 1998, p. 8). Although scholars have not clarified the exact difference between writing development and writing proficiency, a reasonable

assumption is that there would be overlap; nevertheless, the two constructs cannot be considered the same in writing research. As a result, the relationships between NPs and the two constructs (i.e., writing development and writing proficiency) are both included in the discussion below.

Generally, NPs play different roles in exploring writing development and writing proficiency in writing research. In terms of writing proficiency, scholars in applied linguistics have conducted studies to predict writing proficiency based on regression models. The primary research purpose of such studies has been to identify variables (also called "predictors") that can effectively predict writing proficiency. The best predictors in these studies are mostly macro-level variables (e.g., text length and sentence length), which may make fine-grained features (i.e., specific lexico-grammatical features) overshadowed in regression models. Among the fine-grained features, NPs and noun modifiers have rarely been included in any best prediction models in previous studies. For writing development, scholars have conducted studies to distinguish developmental stages. The major purpose of such studies has been to investigate difference(s) in the use of fine-grained features by different groups of students. Specific pedagogical implications have also been provided based on these fine-grained features to facilitate writing development. As a frequent fine-grained feature in advanced academic writing, NPs (and noun modifiers) have been demonstrated to have the potential to distinguish different stages of writing development. In the follow sections, I categorized writing studies on the topic of NPs into (1) predicating writing proficiency and (2) distinguishing writing development. The roles of NPs are summarized and discussed accordingly.

### *4.3.1. Predicting Writing Proficiency with Noun Phrases*

Previous studies have applied a large set of variables to predict writing proficiency[16] based on multiple linear regression in various written genres (e.g., test essays and graduate writing), but NP-related features[17] (e.g., nouns and noun modifiers) have not been included in the best prediction models. However, the exclusion of NPs (and noun modifiers) is not surprising because a large portion of variance in writing proficiency is often explained by macro-level variables (e.g., text length and sentence length). Fine-grained variables (e.g., NPs) are not supposed to be as effective as macro-level variables in predicting writing proficiency. For instance, Ferris (1994) examined 28 lexico-grammatical features in placement test essays (N = 160) written by L2 students to explore which features could effectively predict L2 writing proficiency. Two NP-related features were included in the study, namely, nominal forms and relative clauses. The exam essays were rated by three independent raters on a 10-point rating scale. The 28 features were analyzed by multiple linear regression as predictors, and the best predictors were identified with a stepwise analysis. Ferris (1994) found that the best model included five predictors that explained 50.3% variance of the writing scores (R-square = 0.53): number of words, synonymy/antonymy, word length factor, passives, and 3rd person pronouns. Neither nominal forms nor relative clauses were included in the final regression model. Their exclusion suggests that these two features were not effective predictors for L2 writing proficiency in the placement test essays.

[16]Writing scores are conceptualized using two different terms in the existing literature: "writing quality" and "writing proficiency". As the writing scores are primarily from placement tests and standardized tests that measure writing proficiency, the term "writing proficiency" is used in this dissertation.

[17] The term "NP-related features" refers to grammatical features and measures associated with NPs, including nouns, noun modifiers, nominal forms, number of noun modifiers, etc.

With a growing integration of computational techniques in applied linguistics, researchers began to study a wider range of features in L2 writing research. Guo, Crossley, and McNamara (2013) studied 70 linguistic features related to text length, lexical sophistication, syntactic complexity, and cohesion in a corpus of 240 TOEFL iBT independent essays (mean length = 125-225 words) and 240 TOEFL iBT integrated essays (mean length > 300 words). Three NP-related features were included in the study: nominalizations, number of modifiers per noun phrase, and noun overlap. The essays were rated by trained human raters on a 5-point scale (see information on TOEFL website). Coh-metrix, a computational program for text analysis, was used to calculate the frequencies of the selected features. A multiple linear regression was run with a stepwise analysis to identify the best predictors for the TOEFL writing scores. For the integrated essays, seven predictors were included the best model (R-square=0.73), including number of words per text, past participle verbs, word similarity (content words), verbs in 3rd person singular, semantic similarity, verbs in base form, and word frequency (content words). However, none of the seven predictors were NP-related features. For the independent essays, the best model (R-square = 0.57) consisted of five predictors: number of words per text, average syllables per words, noun hypernym values, past participle verbs, and conditional connectives. The only NP-related feature in the final regression model was noun hypernym values, a semantic measure. Therefore, Guo et al. (2013) also suggests that NP-related features are weak predictors of L2 writing proficiency in the TOEFL writing tasks.

Crossley, Kyle, Allen, Guo, and McNamara (2014) replicated as Guo et al. (2013) with a larger sample size of TOEFL independent essays. Crossley et al. (2014) examined 59 linguistic features associated with multiple linguistic levels (e.g., length, complexity, cohesion, and rhetorical style) to explore their potential to predict L2 writing proficiency. Altogether, 480 rated

TOEFL independent essays were used, and the frequencies of 59 features were provided by Coh-Matrix. Based on a multiple linear regression, Crossley et al. (2014) proposed the best model (R-square = 0.60) to predict the writing scores with six predictors, and the model explained 60% of the variance in human-rated scores of the essays. The six predictors included number of word types, word imageability, proportion of key words, incidence of "and", Latent Semantic Analysis body to conclusion, and incidence of perfect verb forms. Consistent with Guo et al. (2013), no NP-related features were included in the best model, so a similar conclusion can be drawn: the NP-related features do not play a strong role in predicting writing proficiency.

The previous two studies applied a wide range of features in multiple linguistic levels, such as length, cohesion, and complexity, among others. As grammatical complexity has recently received growing research attention in the 2010s, Crossley and McNamara (2014) focused on only using syntactic complexity measures to predict writing proficiency. In their study, 11 clausal and phrasal features were examined in essays (N = 57) written by college L2 students in an English for Academic Purposes (EAP) program. These essays were scored by two expert raters based on five analytical domains (i.e., content, organization, vocabulary, language use, and mechanics). The interrater reliability was 88% for the overall writing scores. A multiple linear regression was run with a stepwise analysis to find the best model (R-square = 0.32) with three predictors on the overall writing scores, meaning that the 32% of the variance of the overall scores could be explained by three predictors (i.e., three syntactic measures) in the regression model. However, no NP-related features were in the model (i.e., number of modifiers per noun phrase and number of relative clauses). Therefore, Crossley and McNamara (2014) reached a similar conclusion that NP-related features cannot effectively predict writing proficiency in EAP essays.

In addition, Yang, Lu, and Weigle (2015) applied syntactic complexity measures to predict writing proficiency in graduate writing. In this study, 14 syntactic complexity measures were investigated in a corpus (124,830 tokens) of 380 graduate essays. The essays were written by L2 graduate students on the topics of personal appearance and future plans. The syntactic measures were all composite measures related to sentence length, coordination, subordination, and phrases. In terms of NPs, the study measured complex NPs per clause and complex NPs per T-unit. To examine the prediction of the syntactic complexity measures on writing scores, a multiple linear regression was run with an all-possible-subsets analysis to identify effective models. Table 4 presents the "best models" suggested by Yang et al. (2015) as well as the models that include NP-related features. These measures in the regression models were also used as complexity measures in this study, although some measures have validity problems for measuring grammatical complexity, for instance, mean length per clause (MLC), which has already been used a fluency measure in previous studies (see Chapter 2 for more information).

Table 4. Regression Models for the Two Topics

| Topic | The MLR Models | Adjusted R-square |
|---|---|---|
| Personal appearance | MLC, DC/TU, TU/S* | 0.07 |
| | MLC, DC/TU, TU/S, CNP/C | 0.07 |
| Future plan | MLC, TU/S, DC/TU* | 0.08 |
| | DC/TU, TU/S, NFE/C, CNP/C | 0.08 |

*Note*. * marks the "best" model proposed in Yang et al. (2015). MLC = mean length of clause; DC/TU = dependent clauses per T-unit; TU/S = T-units per sentence; CP/C = coordinate phrases per clause; CNP/C = complex noun phrases per clause; NFE/C = non-finite elements per clause.

Yang et al. (2015) found that the best model to predict writing scores for the topic of personal appearance consisted of three predictors: mean length per clause, dependent clauses per T-unit, and T-units per clause. The adjusted R-square was 0.07, which means that the three predictors explained 7% of the variance of the writing scores. Another suggested model

consisted of four predictors: mean length per clause, dependent clauses per T-unit, T-units per clause, and complex NPs per clause. The adjusted R-square was also 0.07. By comparing the two models, it can be concluded that including complex NPs per clause in the regression model contributes to explaining little variance of the writing scores for the topic of personal appearance.

According to the future plan topic, the best model also consisted of three predictors: mean length per clause, T-units per clause, and dependent clauses per T-unit. The adjusted R-square was 0.08, suggesting that the three predictors explained 8% of the variance of the writing scores. Another model consisted of four predictors: T-units per clause, dependent clauses per T-unit, non-finite elements per clause, and complex NPs per clause. The adjusted R-square was 0.08 as well. This means that the combination of nonfinite elements per clause and complex NPs per clause made the exact same contribution to predicting the writing scores as dependent clauses per T-unit. This reveals that the mean length per clause has a multicollinearity with the non-finite elements per clause and complex NPs per clause. Therefore, mean length per clause has a similar power on score prediction as the combination of the non-finite elements per clause and complex NPs per clause, so complex NPs per clause were not included in the best model. In this case, it was found complex NPs play a very weak role in predicting writing proficiency in graduate writing.

### *4.3.2. Discriminating Writing Developmental Stages with Noun Phrases*

Writing development has a strong relation with language acquisition, and language learners in different developmental stages often acquire and produce different linguistic features (Crossley & McNamara, 2014). The relation between NPs and writing development has received a growing amount research attention in the 2010s in both cross-sectional and longitudinal research. In cross-sectional research, one of the most influential theoretical frameworks is the

aforementioned developmental index of writing complexity features proposed in Biber et al. (2011), including 11 noun modifiers (see Chapter 3). Parkinson and Musgrave (2014) conducted a corpus-based study to investigate the 11 noun modifiers in L2 essays written by EAP students (13,711 tokens) and MA TESOL students (12,577 tokens). With a Fisher's exact test, it was found that six noun modifiers in the index were significantly different between the two groups. MA TESOL students used significantly more relative clauses, premodifying nouns, prepositional phrases, -ed clauses, and appositive noun phrases, whereas the EAP students used significant more attributive adjectives. Parkinson and Musgrave observed that the EAP students relied on this basic noun modifier (i.e., attributive adjectives), but the MA TESOL students used more advanced noun modifiers (e.g., prepositional phrases, appositive noun phrases) in their essays. This suggests that complex NPs with different noun modifiers have the potential to discriminate different stages of writing development.

Wang and Beckett (2017) studied a set of noun modifiers in personal statements written by Chinese EFL students (25,414 tokens) and advanced English writers (11,822 tokens). The study included 12 noun modifiers taken from both Biber et al.'s (2011) index and Halliday and Matthiessen (2014). The authors ran a Chi-square analysis to test if writing development affected the use of these noun modifiers. The results demonstrated that Chinese EFL students used premodifiers (e.g., determiners, numeratives, and attributive adjectives) more than expected[18], whereas the advanced English writers used postmodifiers in their personal statements more than expected, such as prepositional phrases, clauses as postmodifier, and appositive noun phrases. In the index, all postmodifiers were in the advanced stages of writing development. Therefore,

[18] "Expected" is used as a term to refer to a statistical expected value in a Chi-square test.

Wang and Beckett provided additional empirical evidence that complex NPs can differentiate the stages of writing development within their research context.

The two studies mentioned above used small corpora, but the noun modifiers have also been examined in large-scale corpora across different genres and disciplines by corpus linguists. Staples et al. (2016) investigated 23 lexico-grammatical features, including nouns, nominalizations and seven noun modifiers (i.e., attributive adjectives, premodifying nouns, of-genitives, *that* noun complement clauses, infinitive clauses, relative clauses, and nonfinite relative clauses) in a subset corpus of the British Academic Written Corpus. The corpus (4.5 million tokens) was designed based on four written genres (i.e., essays, critiques, case studies, and explanations) and four disciplines (i.e., arts and humanities, social sciences, life sciences, and physical sciences) across four academic levels (first, second, and fourth year undergraduate students, and graduate students). Staples et al. (2016) applied an ANOVA analysis with a bonferroni adjustment ($p=0.05/23=0.002$) to analyze how academic levels, disciplines, and written genres influenced the use of these lexico-grammatical features. The results revealed that there were significant differences in the frequencies of the nouns, nominalizations, and all seven noun modifiers across the four academic levels, even though the differences were mediated by disciplines and written genres. A post-hoc analysis demonstrated (1) the frequencies of three phrasal NP-related features (i.e., nominalizations, attributive adjectives, and premodifying nouns) displayed a sharp increase from the fourth-year undergraduate to the graduate level; and (2) in contrast, the frequencies of the two clausal noun modifiers (i.e., relative clauses and noun complement clauses) showed a sharp decrease from the fourth-year undergraduate to the graduate level. Biber et al. (2011) claimed that as academic levels become advanced, academic writing skills develops, and writers prefer a nominal writing style based on compressed NPs.

Staples et al. (2016) support this claim and demonstrate that phrasal noun modifiers and writing development are related.

In additional to the research based on Biber et al.'s (2011) index, applied linguists have also used composite measures to investigate NPs and writing development. Lu (2011) examined complex nominals per clause and complex nominals per T-unit in essays written by L2 Chinese students across four undergraduate levels. Based on an ANOVA test, Lu investigated the influence of academic levels with the two composite measures. Significant differences between each of the two adjacent undergraduate levels were found for the two measures, which is consistent with the assumption in Wolfe-Quintero et al. (1998) that complex nominals could discriminate different stages of writing development.

Ravid and Berman (2010) examined the development of NPs in different age groups in L1 writing. They studied 48 native speakers of English who were at different age groups. The 48 participants were divided into four age groups equally, with 12 participants in each group: 9-10 year-olds (middle childhood), 12-13 year-olds (pre-adolescence), 16-17 year-olds (adolescence), and 20-30 year-olds (adulthood). Four different measures of NP complexity were included:

1. Length score (based number of words in NPs)
2. Number of modifiers
3. Syntactic depth score (based on the number of syntactic nodes in NPs)
4. Syntactic variability score (based on a specific coding scheme on NP patterns)

Ravid and Berman (2010) applied an ANOVA test with a post-hoc analysis on these four measures among the four different age groups. The result showed that age had a large effect on the NP measures. First, the length scores in the adult group were significantly longer than the other three groups; second, the number of modifiers was significantly greater in the adolescence

and adult groups than the middle childhood and pre-adolescence groups; third, syntactic depth scores were significantly greater in the adolescence and adult groups than in the middle childhood and pre-adolescence groups; fourth, syntactic variability scores were significantly greater in the adult groups than in the other three groups. Ravid and Berrman's (2010) work thus supports the proposition that NPs can discriminate writing development based on four types of measures: length, number of modifiers, syntactic depth, and syntactic variability.

The relationship between NPs and writing development has not only been explored in the cross-sectional studies but also in some longitudinal studies. Crossley and McNamara (2014) included a longitudinal analysis on the 11 selected clausal and phrasal features in essays written by 57 college L2 students at the beginning and end of an EAP program. Crossley and McNamara used a one-way ANOVA test to explore the 11 selected features in the essays. In terms of NPs, the findings demonstrated there was a significant difference in the total number of modifiers per NP between the essays written at the two time periods. The essays written at the end of the semester included more noun modifiers than the essays written at the beginning of the semester. To interpret this finding, Crossley and McNamara (2014) summarized that "L2 learner growth was associated with nominal style and phrasal complexity" (p. 76).

Bulté and Housen (2014) conducted a similar study that examined short-term development in L2 writing complexity. Ten lexical and syntactic measures were examined in the essays written by 45 ESL learners at the beginning and the end of a four-month EAP program, including a measure of phrasal complexity (i.e., mean length of NPs). Paired-sample t-tests were run to examine the difference of the ten measures in the two drafts of the essays. The results showed that six out of ten measures were significantly different. The mean length of NPs was significantly longer in the essays written at the end of the EAP program than at the beginning.

They argued that the six measures can capture the grammatical changes in L2 writing in a short-term EAP program. Also, this study adds empirical support that NPs can distinguish short-term writing development based on length.

Mazgutova and Kormos (2015) explored the change of lexico-grammatical complexity in L2 academic writing over a one-month EAP program at a university in the UK. L2 students were put in a low-proficiency group (5.9 < IELTS writing score < 6.3) and a high-proficiency group (6.4< IELTS writing score < 6.7). The texts in this study were argumentative essays written at the beginning and at the end of the EAP program. The NP-related measures were modifiers per NP, complex nominals, simple postmodifiers (i.e., one postmodifier in a NP), complex postmodifiers (i.e., more than one postmodifiers in a NP), and relative clauses. A Wilcoxon signed-rank test was run to test the differences between the two essay drafts in the low-proficiency and high-proficiency groups. The results revealed that the low-proficiency group produced significantly more features related to complex NPs (i.e., total number of modifiers, complex nominals, relative clauses, and complex postmodifiers) between their two essays, whereas the high-proficiency group did not produce significantly more of these features between their two essays. As a result of their research, Mazgutova and Kormos argued that L2 students with low writing proficiency have great potential to enhance their use of NPs via an EAP program, where they are exposed to academic reading, get feedback on their writing assignments, and are immersed in an L1 environment.

Staples and Reppen (2016) conducted a lexico-grammatical analysis across three variables—L1s (English, Chinese, and Arabic), genres (argumentative paper and rhetorical analysis), and language ratings—with a corpus 120 papers written by students in first-year writing courses. The eight lexico-grammatical features analyzed in the study include three noun

modifiers (i.e., attributive adjectives, premodifying nouns, and noun complement clauses) and five other features (i.e., verb complement clauses, adverbial clauses with causative, conditional and other functions, and type/taken ratio). The results show, from the perspective of writing development, important similarities in the use of lexico-grammatical features among writers from the three L1 groups. Surprisingly, L2 (Chinese) students produced premodifying nouns significantly more than L1 students, which is related to advanced level writing. Based on a qualitative analysis of adjective-noun and noun-noun sequences, Staples and Reppen found that there was more repetition of premodifying nouns and attributive adjectives in L2 writing than in L1 writing. They then argue that "we should be mindful that greater use does not always mean greater development" (p. 30).

Yoon and Polio (2017) investigated how time and genre influenced writing development. Complex nominals per clause and complex nominals per T-unit were examined with Lu's L2 Syntactic Complexity Analyzer in the writing of 37 ESL students over a semester (15 weeks). The writing consisted of argumentative and narrative essays written during six different times (with a three-week interval). With an ANOVA analysis, Yoon and Polio found that genre had a significant effect on the two measures of complex nominals, whereas time had no significant effect. This study failed to support that NPs could distinguish writing development over a certain period of time. At the end of the study, Yoon and Polio argued that one semester was a short period of time for clear grammatical growth and that differences could be examined from a longer time interval. Overall, however, most of the empirical studies introduced above (both cross-sectional and longitudinal) suggest that NPs are associated with different stages of writing development.

### *4.3.3. Research Prototype*

Lan, Lucas and Sun (2019) is a recent study to analyze NP complexity in L2 writing. To the best of my knowledge, this study is the first one to apply a Chi-square test with a residual analysis to study NP complexity (grammatical complexity in general) in applied linguistics. I discuss it separately from others because my dissertation can be considered a partial replication study of Lan, Lucas and Sun (2019). Wang and Beckett (2017) applied a *Chi-Square Test*, but it was not followed by a *Residual Analysis*. In their study, a *One-way Chi-square Test* was applied 21 times on 13 different noun modifiers (e.g., classifier-adjectives, clause as modifiers, and appositives) and eight part-of-speech features related to NPs (e.g., nouns, pronouns, adjectives, and determiners). It is inefficient to use the same statistical test a number of times in empirical studies, and a *Residual Analysis* effectively avoids the repeated use of a *One-way Chi-square Test*.

Lan, Lucas and Sun (2019) investigate how NP complexity is influenced by L2 writing proficiency based on the 11 noun modifiers in Biber et al.'s (2011) index. Their corpus is a subset (i.e., 100 argumentative papers written by L2 students) of Corpus and Repository of Writing (Crow)[19]. Fifty papers are from high-proficiency students (TOEFL writing score > 23), and 50 papers are from low-proficiency students (18 < TOEFL writing score < 23). Using a Chi-square test on the normed frequencies of the 11 noun modifiers, Lan, Lucas and Sun found a significant association between the noun modifiers and the students' writing proficiency levels. Then, via a residual analysis, they identified four noun modifiers that contribute to the association the most, namely attributive adjectives, relative clauses, premodifying nouns, and prepositional phrases (of). Based on a qualitative analysis of the argumentative papers, Lan,

[19] Learn Corpus and Repository of Writing at https://writecrow.org/

Lucas and Sun interpret the quantitative findings and also provide suggestions for different ways of teaching these four noun modifiers in writing courses. In the dissertation, I follow the statistical methods used in Lan, Lucas and Sun (2019) (i.e., a Chi-square test followed by a residual analysis), which is not often applied in writing research on grammatical complexity.

### 4.4. Research Gaps and Research Questions

NP complexity is a fairly recent research topic in L2 and academic writing. Most relevant studies were conducted in the 2010s. These studies have found: (1) compressed NPs are frequent in academic writing; (2) NP-related features (i.e., noun, nominalizations, noun modifiers, and composite NP measures) cannot effectively predict L2 writing proficiency; (3) NP-related features can effectively distinguish different stages of writing development. After reviewing these studies, I decided to follow the third research trend to explore how NP complexity can serve as a discriminator for writing development in academic writing, which is promising and sustainable.

I identified two gaps in the existing literature that are worth being bridged. First, previous studies have primarily investigated the relationship between NP complexity and writing development primarily in L2 writing. Few studies include L1- and L2-English language background as a variable to see how it influences NP complexity in academic writing. Second, NP complexity has been primarily studied with tests of significance and regression. Few studies have applied statistical tests from other perspectives. To bridge the gaps, L1- and L2-English language background is included as a variable in this dissertation along with academic levels (i.e., undergraduate and graduate), which suggests two common stages of writing development. Also, I replicate the statistical method used in Lan, Lucas and Sun (2019) to examine the effects

of academic level and L1- and L2-English language background on NP complexity in my corpus (i.e., specific writing samples) from a new statistical perspective (i.e., probability).

This dissertation investigates how NP complexity is influenced by academic levels and L1- and L2-English language background in academic writing. The operationalization of the research constructs is: (1) NP complexity is operationalized as 11 noun modifiers in Biber et al.'s (2011) index; (2) Academic level is operationalized to undergraduate and graduate levels; and (3) L1- and L2-English language background is operationalized to L1 and L2. The dissertation includes two sets of research questions to be answered (six questions in total):

Question Set A

1. Is there an association between academic level and NP complexity in academic writing?
2. What is the strength of association?
3. Which specific noun modifiers contribute to this association the most?

Question Set B

4. Is there an association between L1- and L2-English language background and NP complexity in academic writing?
5. What is the strength of association?
6. Which specific noun modifiers contribute to this association the most?

# CHAPTER 5. **METHODS**

## 5.1. Corpus-based Approach and Register Analysis

A *Corpus-based Approach* is applied in this dissertation. A corpus-based study is described as one in which "corpus linguistic researchers are guided by previous corpus findings or by specific issues concerning language use. That is, researchers have a very specific idea before searching the corpus as to what linguistic item they are looking for in a corpus" (Crawford & Csomay, p. 10-11). An important characteristic of the corpus-based approach is to examine distributional differences of a set of pre-selected grammatical features across different situations of language use (Biber & Conrad, 2009). The method used in this dissertation is consistent with the characteristics of the corpus-based approach. My dissertation examines NP complexity based on the 11 noun modifiers proposed in Biber et al.'s (2011) hypothesized developmental index of writing complexity and analyzes how the use of these noun modifiers is associated with two variables that represent situations of language use (i.e., academic writing, L1- and L2-English language background).

The core spirit of register analysis is that grammatical forms have different grammatical functions in different situations of language use (i.e., registers). "[People] choose to use particular linguistic features because those forms fit the communicative context of the text, whether in conversation, a political speech, a newspaper editorial, or an academic research article" (Biber & Conrad, 2009, p. 3). The framework of register analysis proposed by Biber and Conrad (2009) consists of a situational analysis, a linguistic analysis, and a functional interpretation. The major quantitative findings of this dissertation were interpreted based on the grammatical functions of noun modifiers in academic writing, and the study focuses on linguistic analyses; however, a functional interpretation is a secondary focus.

## 5.2. Corpus

The corpus of this dissertation is a subset of the British Academic Written English Corpus (BAWE). BAWE represents academic writing from university students in the United Kingdom (UK). In total, the BAWE corpus contains 2,761 files, about 6,506,996 words. BAWE includes texts written by both L1 and L2 university students from four academic levels (i.e., first-year, second-year, and fourth-year undergraduate students, and graduate students) across four disciplinary areas (i.e., Arts and Humanities, Social Sciences, Life Sciences, and Physical Sciences) and across 13 academic written genres in universities (e.g., case studies, critiques, research reports, and essays). BAWE is a publicly available corpus and the files in BAWE are freely downloadable from the Coventry University website[20]. An Excel sheet is provided along with the files to show the metadata of the files in BAWE, such as disciplines, genres, length, L1 background, and academic level, among others.

The corpus in this dissertation was built by extracting 200 files from BAWE. Staples et al. (2016) reported empirical evidence showing the influence of academic disciplines and written genres on a number of lexico-grammatical features (e.g., noun modifiers) in academic writing. Therefore, considering the existing data in BAWE, I selected essays (a specific and common written genre in the universities) written by students in the social sciences and humanities in the corpus of this dissertation. With respect to the two variables investigated in the dissertation, the academic levels of the students are categorized into undergraduate and graduate, whereas the L1- and L2-English language background of the students are categorized into L1 and L2. *Topic* is not controlled in the dissertation, because BAWE only offers academic writing on different topics.

[20]https://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/

For a large-scale corpus-based study, it is difficult to find or create a large collection of files on the same topic. Therefore, different files have different topics in the corpus of this dissertation.

Table 5. Basic Information of the Corpus in this Dissertation

| | L1 Undergraduate | L1 Graduate | L2 Undergraduate | L2 Graduate | Main Corpus |
|---|---|---|---|---|---|
| File no. | 50 | 50 | 50 | 50 | 200 |
| Tokens | 108,828 | 189,169 | 117,989 | 198,855 | 615,841 |
| Mean Length | 2196.56 | 3783.38 | 2359.78 | 3977.10 | 3079.21 |
| | Undergraduate | Graduate | L1 | L2 | Main Corpus |
| File no. | 100 | 100 | 100 | 100 | 200 |
| Tokens | 226,817 | 388,024 | 297,997 | 316,844 | 615,841 |
| Mean Length | 2,268.17 | 3,880.24 | 2,979.97 | 3,168.44 | 3079.21 |

Table 5 illustrates the design of the main corpus (N = 200). It includes four subcorpora labeled as L1-Undergraduate (N = 50), L1-Graduate (N = 50), L2-Undergraduate (N = 50), and L2-Graduate (N = 50). To be specific, the total tokens for the main corpus is 615,841, and total tokens for each subcorpora is 108,828 (L1-undergraduate), 189,169 (L1-graduate), 117,989 (L2-undergraduate), and 198,855 (L2-graduate). The mean length of the files in the main corpus is 3079.21, and the mean length of each subcorpus is 2196.56 (L1-undergraduate), 3783.38 (L1-graduate), 2359.78 (L2-undergraduate), and 3977.10 (L2-graduate).

The four subcorpora demonstrate a balanced design of the main corpus, and they are combined to build the corpora related to the two variables (i.e., academic level and L1- and L2-English language background). For instance, the L1 undergraduate and L2 undergraduate corpora were combined to build the undergraduate corpus. To be more specific, (1) the undergraduate corpus (N = 100) has 226,817 tokens and the mean length of the files is 2,268.17;

(2) the graduate corpus (N = 100) has 388,024 tokens and the mean length of the files is 3,880.24; (3) the L1 corpus (N = 100) has 297,997 tokens and the mean length of the files is 2,979.97; (4) the L2 corpus (N = 100) has 316,844 tokens and the mean length of the files is 3,168.44.

## 5.3. Instruments

### *5.3.1. Biber Tagger*

The corpus of this dissertation was tagged by the *Biber Tagger*, which was developed by Douglas Biber from Northern Arizona University. The tagger can annotate part-of-speech (POS) information and a number of additional lexico-grammatical features, which is critical for this dissertation, such as relative clauses, noun complement clauses, and participial clauses. The *precision* [21] and *recall* [22] of the *Biber Tagger* are close to 95% for L1 and are over 90% for L2 writing (Biber & Gray, 2013). Other studies that have applied the tagger have also reported high precision and recall. For instance, Staples et al. (2016) calculated the accuracy rate for 23 lexico-grammatical features in their study and found that "the final recall and precision rates all approached 90% or better" in L1 writing" (p. 159). With respect to NP complexity, Lan, Lucas and Sun (2019) reported that the precision rates of the noun modifiers that can be tagged were approaching 90% in L2 writing, including 99.70% for attributive adjectives, 93.86% for relative clauses, and 88.67% for noun complement clauses. I present and discuss the precision and recall related to the 11 noun modifiers in this chapter (see Table 6).

[21] Precision refers to the fraction of correct tags among all the tags of a grammatical feature.
[22] Recall refers to the fraction of annotated tags among all actual cases of a grammatical feature.

### *5.3.2. Python Programs*

I built four computer programs with Python 3.0 to extract the noun modifiers from the tagged corpus and to calculate the raw frequencies of the noun modifiers. Regarding the four programs, Program A was built to extract the 10 noun modifiers from the tagged corpus and to output the extracted noun modifiers in concordance-line format. Appositive NPs are not tagged by the Biber Tagger, so they were extracted from the untagged corpus rather than the tagged corpus. It is important to mention that, among the 10 noun modifiers, Program A only extracts (1) relative clauses with relativizers (e.g., *which*, *who*, *that*) (2) noun complement clauses with *that* as complementizer (e.g., *the fact that grammatical complexity is important for writing studies*). Program B was built to categorize all the concordance lines of the noun modifiers based on their types, which is convenient for making manual adjustments of the concordance lines. Program C was built to count the raw frequencies of the 10 noun modifiers and output the frequencies in a tab-separated format, which can be imported directly into statistical packages. Finally, program D was built to extract appositive NPs from the untagged corpus based on regular expressions for specific patterns. More details regarding the application of the instruments are shared in the next section (5.4 Procedures).

## 5.4. Procedures

This section explains the four main procedures in the dissertation:

1. Corpus tagging and tagging evaluation
2. Automatic extraction with the Python programs
3. Manual adjustment and frequency counting
4. Statistical Analysis

### *5.4.1. Corpus Tagging and Tagging Evaluation*

The corpus in this dissertation was firstly tagged by the *Biber Tagger* (see Appendix A for an excerpt of a tagged file). Next, the tagged corpus was evaluated for tagging precision and recall. Ten files (5% of the total files) were sampled for tag checking, and two trained human coders (i.e., two PhD candidates) assigned gold labels[23] to eight lexico-grammatical features that are closely related to the noun modifiers: nouns, attributive adjectives, relative clauses, prepositional phrases, -ing clauses as post-noun modifiers, -ed clauses as post-noun modifiers, noun complement clauses, and infinitive clauses. Intercoder reliability[24] of the gold labels reached 98.6% in a training session, and the two human coders then assigned gold labels individually on the sampled files.

The two human coders then evaluated the tags of the eight target features based on the gold labels. Precision and recall were calculated based on four specific relationships between the gold labels and the tags of the target features: (1) *true positives* (2) *false negatives* (3) *false positives* and (4) *true negatives* (Pustejovsky & Stubbs, 2012). Next, the formula in Pustejovsky and Stubbs (2012) was used to calculate the precision and recall of the eight lexico-grammatical features (see Appendix B for example).

**Formula**

1. Precision = true positive / (true positive + false positive)
2. Recall = true positive / (true positive + false negative)

Table 6 presents the precision and recall for the target features. Five target features had both precision and recall approaching or greater than 90.00%, namely, nouns, attributive adjectives,

[23] "Gold labels" refer to the labels assigned by human coders, which is considered correct labels.
[24] Intercoder reliability refers to agreement of the assigned gold labels between the two human coders.

relative clauses, prepositional phrases, and infinitive clauses. In contrast, -ing clauses as noun modifiers and -ed clauses as noun modifiers had low precision (85.71%, 76.00%) and recall rates (77.42%, 51.35%). For noun complement clauses, the recall was 85.00%, which can be considered acceptable but still lower than the expected 90.00%.

Table 6. The Precision and Recall of the Biber Tagger on the Corpus

| Target features | Precision | Recall |
|---|---|---|
| Nouns | 95.88% | 99.05% |
| Attributive adjectives | 89.57% | 93.21% |
| Relative clauses | 97.60% | 96.54% |
| Prepositional phrases | 98.05% | 94.24% |
| -ing clauses as modifiers* | 85.71% | 77.42% |
| -ed clauses as modifiers * | 76.00% | 51.35% |
| Noun complement clauses (that)* | 89.47% | 85.00% |
| Infinitive clauses | 94.24% | 98.42% |

*Note*. The features that needs to be fixed are labelled with asterisks, and their precision and recall rates are shaded in grey. Noun complement clauses is checked for a specific category, which is led by the complementizer *that*.

Last, five human coders (i.e., five PhD candidates) with backgrounds in applied linguistics qualitatively fixed the tags of -ing clauses as noun modifiers, -ed clauses as noun modifiers, and noun complement clauses (that) in all the 200 essays in my corpus. During the training session, the intercoder reliability[25] was 96.9%. The human coders then fixed the corresponding tags individually. By the end of this step, the tagged corpus was ready to be used.

### *5.4.2. Automatic Extraction with the Python Programs*

The extraction of the 11 noun modifiers consisted of automatic extraction and manual adjustment. For the automatic extraction, the Python programs were applied to extract the 11

[25] Intercoder reliability refers to agreement of the fixed tags among the five human coders on -ing clauses as noun modifiers, -ed clauses as noun modifiers, and noun complement clauses.

noun modifiers in Biber et al.'s (2011) index of writing complexity features. The automatic extraction was based on two specific categories (Lan & Sun, 2019):

1. Direct extraction based on the tags: the grammatical functions of five noun modifiers can be tagged by the *Biber Tagger*: attributive adjectives, relative clauses, -ing clauses. -ed clauses, and noun complement clauses (that). Therefore, these noun modifiers were directly extracted from the corpus based on the tags.

2. Indirect extraction based on chunking patterns: the grammatical functions of the other six noun modifiers are not directly tagged by the tagger, namely, premodifying nouns, prepositional phrases (of), prepositional phrases (other), infinitive clauses, prepositional phrases followed by -ing clauses, and appositive noun phrases. As a result, chunking patterns were used for extraction. Chunking refers to grouping words into meaning linguistic chunks (Bird, Klein, & Loper, 2009). The chunking patterns for these six noun modifiers are: (1) *noun + noun* for premodifying nouns; (2*) noun + of* for prepositional phrases (of); (3) *noun + prepositions* (except for of) for prepositional phrases (other); (4) *noun + infinitive marker-to* for infinitive clauses as noun modifiers; (5) specific patterns based on parentheses, commas, and dash(es) for appositive NPs (Lan & Sun, 2019).

Example 6 below illustrates the extracted noun modifier in a concordance-line format. The noun modifier (i.e., the relativizer of relative clauses) is highlighted with angled brackets, and 10 words before and after the noun modifier are also printed. Also, the filename and total number of tokens are provided on the top of each file.

**Example 6**

The filename is: 0001a_tagged

The total token is: 774

NOUN_REL: ... mix green tea with boiled water <<<whose>>> temperature is too high ...

It is important to mention an exception, appositive NPs. The extraction of this particular noun modifier is based on the application of regular expressions in the untagged corpus. Two specific chunking patterns were found to be highly frequent in academic writing in Biber et al. (1999): first, noun phrases set off by a pair of commas, as in "*a useful programming language, Python, will be taught*"; second, noun phrases in parentheses, as in "*Corpus and Repository of Writing (Crow)*". I also performed a qualitative check on five randomly sampled files in the

corpus and found another pattern, namely, noun phrases separately by dash(es), as in "*two statistical methods — regression and ANOVA — are both important*". Considering the potential pattern varieties, I also included two additional patterns in my Python programs: (1) an appositive NP separated by dash(es) and a comma as in "*two statistical methods — regression and ANOVA,*" and (2) an appositive NP separated by dash(es) and a period "*two statistical methods — regression and ANOVA*". Precision and recall of appositive NPs are not reported in this section.

### *5.4.3. Manual Adjustment and Frequency Counting*

The noun modifiers based on direct extraction needed no manual adjustment. As the corresponding tags had already been fixed, the cases of these noun modifiers were found over 90% accurate. These noun modifiers included attributive adjectives, relative clauses, -ing clauses. -ed clauses, and noun complement clauses. In contrast, the cases of noun modifiers based on indirect extraction were manually adjusted because chunking patterns cannot guarantee the accuracy of the extracted cases. For example, I used the noun-preposition sequence to extract all potential cases of prepositional phrases as postmodifiers. Although this chunking pattern does not work with a high level of accuracy, it was the most feasible pattern that could be used for automatic extraction.

The noun modifiers that needed manual adjustments were premodifying nouns, prepositional phrases (of), prepositional phrases (other), infinitive clauses, and prepositional phrases followed by -ing clauses. Three human coders (i.e., three PhD candidates) qualitatively read the concordance lines for each noun modifier to exclude incorrect cases from the files. During the training session, the intercoder reliability was 90.25%, so the human coders adjusted these noun modifiers individually (see Appendix C for the training materials).

By the end of this step, the extracted cases of all the noun modifiers were considered as sufficiently accurate (accuracy rate > 90%) for statistical analysis. Next, based on the adjusted files, I applied my program to count the frequencies of the 11 noun modifiers one by one. The raw frequencies were output into a plaintext file with the tab serving as a delimiter to make these files importable into *Statistical Package for Social Sciences 23* (SPSS) for quantitative analysis.

### *5.4.4. Statistical Analysis*

The statistical analysis in this dissertation is a replication of the statistical methods in Lan, Lucas and Sun (2019). Two statistical analyses were run to address the research questions, *Chi-square test* and *residual analysis*. First, two datasets were built in SPSS for the two variables (i.e., academic level, L1- and L2-English language background). The first dataset was based the normed frequencies of the 11 noun modifiers (per 100,000 words) in the undergraduate corpus and the graduate corpus, respectively, and labels were added in the dataset to demonstrate the academic levels (i.e., undergraduate, graduate) in the dataset. Similarly, the second dataset was based on the normed frequencies of the 11 noun modifiers (per 100,000 words) in the L1 corpus and the L2 corpus, respectively, and labels were also added in the dataset to demonstrate L1- and L2-English language background (i.e., L1, L2). As the size of the main corpus in this dissertation is 615,841 tokens, I normalized the raw frequencies of the 11 noun modifiers to 100,000 tokens to eliminate the influence of different corpus sizes.

Second, two Chi-square tests were run in SPSS with the two datasets to explore: the association between academic level and noun modifiers and the association between L1- and L2-English language background and the noun modifiers, respectively. SPSS offers the options of calculating effect sizes and residuals together with Chi-square values. (1) As the Chi-square tests only detect the significance of association, effect size (e.g., Cramer's V) was calculated to

demonstrate how strong the associations were between academic level and the noun modifiers, and between L1- and L2-English language background and the noun modifiers. (2) A residual analysis was conducted as a post-hoc analysis to demonstrate which specific cells[26] in the contingency tables of the Chi-square tests led to the significance of the associations. The residual analysis in this step is based on adjusted standardized residuals instead of raw residuals to eliminate the influence of cell size. The following paragraphs provides additional information on the statistical analysis.

Three reasons support the decision to use a Chi-square test — a nonparametric test — rather than a parametric test (e.g., ANVOA or a t-test). First, the distributions of six of the 11 noun modifiers violated the requirement of normality for parametric tests. The six modifiers did not display normal distributions based on their Q-Q plots output from SPSS (see Appendix D for the plots), namely attributive adjectives, –ed clauses, prepositional phrases followed by –ing clauses, noun complement clauses, infinitive clauses, and appositive NPs. Second, the noun modifiers violate the requirement of independence for a parametric test. Lexico-grammatical features are not independent, because they are combined together in a text based on certain grammatical rules. The increased use of one feature (e.g., nouns) may generate increased use of another feature (e.g., attributive adjectives) or decreased use of another feature (e.g., pronouns). Third, in terms of nonparametric tests, the Chi-square distribution is robust to different types of distribution, which fits the need of this dissertation.

The combination of a Chi-square test followed by a residual analysis has not been used to explore grammatical complexity in writing research, except for Lan, Lucas and Sun (2019). It is reasonable to assume that this statistical combination has been underrepresented in (L2) writing

[26] In this dissertation "cell" refers to specific levels of the nominal variables, for instance, prepositional phrases (of) used by graduate students or noun complement clauses used by L1 students.

research, and I want offer an alternative method to study grammatical complexity in writing research. To do this, it is necessary to discuss the theory behind this statistical method and the reason why a residual analysis was selected as a post-hoc test. The Chi-square test is used to test "if frequencies of one variable (the dependent variable) change with levels of another independent variable" (Hatch & Lazaraton, 1991, p. 399). The observed frequencies of a dependent variable are tested based on corresponding expected frequencies in order to explore how likely it is that the distribution of the observed frequencies is due to chance. In this dissertation, I tested if the (normed) frequencies of the noun modifiers change with academic level and L1- and L2-English language background, respectively.

A Chi-square test is based on *within-cell calculation*. Table 7 shows a sample contingency table based on the 11 noun modifiers and academic level. To calculate the Chi-square value for this contingency table, we need to follow the formula in Figure 3: $O_i$ stands for an observed value, and $E_i$ stands for a corresponding expected value. In Table 7, an observed frequency (e.g., 80 for attributive adjectives) is compared with the expected frequencies (e.g., 60 for attributive adjectives) within the same cell. In addition, a Chi-square value is a holistic value. For example, the Chi-square value for Table 7 indicates if there is a significant association between the noun modifiers and academic level, but it cannot pinpoint which specific cell(s) contributes to the association the most, such as attributive adjectives used by undergraduate students and/or premodifying nouns used by graduate students. Therefore, a post-hoc analysis needs to be run to identify the source of significance.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Figure 3. This figure shows the formula of Chi-square for test of association

In this dissertation, *residual analysis* is applied as a post-hoc test to identify the cells that make great contributions to a Chi-square value. A residual, as denoted by $O_i$ - $E_i$ in the formula, refers to the difference between observed and expected frequencies. For example, in Table 7, the residual for attributive adjectives used by undergraduate students is 20 (i.e., 80 – 60 = 20). The larger a residual is, the greater a specific noun modifier (e.g., attributive adjectives) used by a specific group (e.g., undergraduate students) contributes to the association.

Table 7. Sample Contingency Table

| Noun Modifiers | Undergraduate | Graduate | Row Marginal |
|---|---|---|---|
| attributive adjectives | Obs: 80<br>Exp: 60 | Obs: 40<br>Exp: 90 | 120 |
| premodifying nouns | Obs: 20<br>Exp: 40 | Obs: 110<br>Exp: 60 | 80 |
| Column Marginal | 100 | 150 | 200 (Table total) |

*Note*. A contingency table includes all the levels of dependent and independent variables of a Chi-square analysis and the observed and expected frequencies at each level. "Obs" refers to observed frequencies and "Exp" refers to expected frequencies. The shaded cell is used as examples in this section. Expected frequency of a cell is calculated as *(row total * column total)/table total*

$$Adjusted\ Standardized\ Residual = \frac{O - E}{\sqrt{E(1 - RowMarginal/n)(1 - ColumnMarginal/n)}}$$

Figure 4. This figure shows the formula of adjusted standardized residuals for in a Chi-square test. "n" refers to table total. O refers to observed frequencies and E refers to expected frequencies. Square root of E refers to an estimate of standard deviation for a specific cell, and the denominator refers to the standard error for a specific cell in a contingency table.

For the residual analysis, it is important to point out that adjusted standardized residuals were calculated in SPSS instead of the raw residual. An adjusted standardized residual is similar to a standardized score (i.e., Z score) in statistics. In this dissertation, adjusted standardized residuals represent the difference between the observed frequencies of the noun modifiers and their corresponding expected frequencies, measured in standard-error units. The formula in

Figure 4 demonstrates: the adjusted standardized residual for a cell in a contingency table is calculated by dividing the raw residual (i.e., observed frequency - expected frequency) by the square root of its corresponding standard error (Sharpe, 2015, p.3), which is estimated as the denominator. For example, in Table 7, the residual for attributive adjectives used by undergraduate students is 20 (i.e., 80 – 60 = 20). Based on the formula shown in Figure 4, the standardized residual is calculated as: $(80 - 60) / \sqrt{60 * (1 - \frac{120}{200})(1 - \frac{100}{200})} = 5.76$.

No *Bonferroni* adjustment was made for the residual analysis. "[I]f the magnitude of the residuals merely serves as a guide to what cells might be of interest, then arguably no adjustment is necessary" (Sharpe, 2015, p. 3). The purpose of running a residual analysis was to explore which noun modifiers used by which group of students contributed to the overall associations the most rather than testing the significance of the adjusted standardized residuals, so it is not necessary for an adjustment. Moreover, the way to decide a cut-off value for large adjusted standardized residuals has not been discussed in second language studies, to the best of my knowledge. Lan, Lucas and Sun (2019) considered the greatest four adjusted standardized residuals as the large residuals in their study. In this dissertation, I applied a cut-off value (|2.00|) to decide the large adjusted standardized residuals in an objective way, which is based on the rule of thumb called "+/- 2 Criteria" in statistics (Sharpe, 2015). As residuals follow a normal distribution, an observed value is greater/less than about 95.4% of other observed values by chance if it is two standard deviations away from the mean. Therefore, I considered an adjusted standardized residual as large if it was less than -2.00 or greater than 2.00 in my dissertation.

*Residual analysis* provides a new angle to think about *frequency* of grammatical features. Based on the description of the calculation process of adjusted standardized residuals, it can be concluded that *residual analysis* takes both observed and expected frequencies into

consideration. This statistical method allows researchers to explore whether high frequencies of grammatical features tend to be meaningful or not. For example, prepositional phrases (of) are the second most frequent noun modifier in my corpus (33.40 per 1,000 words). However, the adjusted standardized residuals, in this study, are small for the cells of prepositional phrases (of) used by undergraduate students (0.0) and used by graduate students (0.0). The high frequency and small adjusted standardized residuals indicate that prepositional phrases (of) are frequent in my corpus only because this noun modifier is expected to be frequent by chance. It is not surprising to see that prepositional phrases (of), which have a wide range of functions in English, tend to be frequent in writing. However, we cannot reveal such information only based on comparing observed frequencies between different groups of users (e.g., graduate students, undergraduate students). More information of prepositional phrases (of) are discussed in Chapter 6 (see 6.4). Overall, *residual analysis* could help researchers analyze *frequency* from a new angle. In L2 writing, scholars have not yet integrated expected frequencies into the analysis of grammatical features. This statistical method is worth of being explored and discussed further in writing studies.

In empirical studies, effect sizes are often reported as a routine to (1) present magnitude of a phenomenon between variables (e.g., the association between the 11 noun modifiers and academic level); (2) indicate if it is meaningful to apply a phenomenon in a practical context (e.g., teaching these noun modifiers in a writing course); (3) suggest the portion that one variable can be explained by the other variable (e.g., how much the writing development between undergraduate and graduate students can be explained by the use of the 11 noun modifiers) (Prentice & Miller, 1992). An effect size is closely related to selection and operationalization of variables in studies (Prentice & Miller, 1992). In this dissertation, I investigate a subcomponent

of language complexity (i.e., NP complexity) and operationalize NP complexity to a small set of grammatical features (i.e., the 11 noun modifiers). A mathematically large effect size is not expected because it is not possible for only the 11 noun modifiers to explain a large portion of associations with academic level and L1- and L2-English language background. Moreover, a mathematically large effect size is often achieved by a cumulative effect of a number of variables. For instance, the effect size between writing proficiency and academic level can be large, because writing proficiency includes a number of variables (e.g., complexity, accuracy, fluency, organization, content, coherence). In this dissertation, there is no strong cumulative effect among the 11 noun modifiers as for writing proficiency. We might not expect a mathematically large effect size as the statistical convention defines.

Cramer's V is reported to show the strength of association (a type of effect size) for the statistical analysis in my dissertation. According to Cohen (1988), there is a convention on interpreting Cramer's V for what values can be referred as little if any association ($V < 0.1$), weak ( $0.1 < V < 0.3$), moderate ($0.3 < V < 0.5$) and strong ($V > 0.5$). As this dissertation is a fine-grained investigation with regard to the 11 noun modifiers, an expectation to achieve a mathematically large effect size is not reasonable. Therefore, a relative perspective to interpret *effect size* (i.e., Cramer's V) is applied in the Results and Discussion, rather than an absolute perspective based on the convention in statistics. Lan, Lucas and Sun (2019) provide an example to interpret Cramer's V from the relative perspective:

- For this fine-grained study, the small effect size (i.e., Cramer's $V = 0.043$) should receive our attention. It is impressive that a small set of grammatical features (i.e., just 11 noun modifiers) can explain approximately 4.3% of the variance of L2 writing proficiency.

# CHAPTER 6. **RESULTS AND DISCUSSION**

In this chapter, I present the results of statistical analysis and discuss the major findings based on these results. Four sections are included in this chapter. In the first section, I present descriptive statistics for the 11 noun modifiers in the corpus and categorize them into different groups based on their normed frequencies. The remaining sections demonstrate the results of *Chi-square test* and *residual analysis* for the two sets of research questions with regard to the association between the 11 noun modifiers and academic level, and the association between the 11 noun modifiers and L1- and L2-English language background, respectively. To appropriately interpret the statistical results, the connections between the major findings and previous literature are discussed, and specific examples from the corpus are provided for a qualitative analysis of the noun modifiers.

## 6.1. Descriptive Statistics

In this dissertation, the Chi-square test, a nonparametric test was selected, and no inferential relation between the sample and population is argued. The results in this dissertation are sample specific, only restricted to the 200 essays in the corpus. In terms of descriptive statistics, I provide the normed frequencies and cumulative percentages of the 11 noun modifiers in the corpus (200 essays). The raw frequencies of the 11 noun modifiers were normalized to 1,000 words to provide a better sense of how these noun modifiers were used. The total normed frequency of the 11 noun modifiers is 152.72, meaning that there are 152.72 noun modifiers per 1,000 words in the main corpus. Figure 5 shows the normed frequencies of the individual noun modifiers and their cumulative percentages. In the corpus, attributive adjectives (64.22) are the most frequent noun modifier. Another four modifiers can be also considered frequent in the

corpus, namely, prepositional phrases (of) (33.40), premodifying nouns (20.64), prepositional phrases (other) (17.27), and relative clauses (7.57). In contrast, the remaining six noun modifiers are infrequent because their normed frequencies are either close to or below 2.00, for instance -ing clauses (1.43) and appositive NPs (1.50).

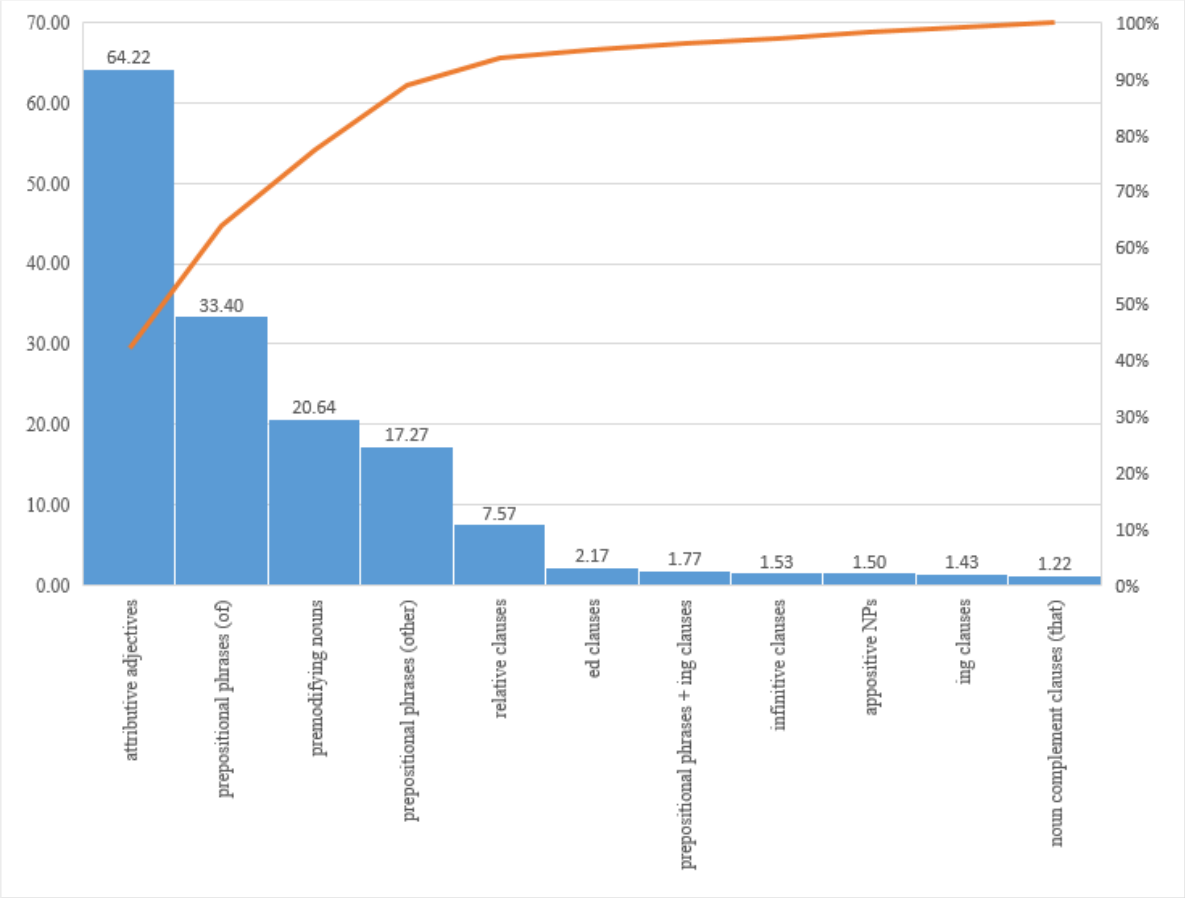



Figure 5. This figure demonstrates the normed frequencies and cumulative percentages of the 11 noun modifiers.

The normed frequencies of the 11 noun modifiers in my corpus are close to the published normed frequencies (per 1,000 words) in academic prose reported in Biber and Gray (2011) and Biber et al. (1999)[27]. In their studies, the frequent noun modifiers are also attributive adjectives

[27] The normed frequencies are firstly summarized in Parkinson and Musgrave (2014) based on the estimation of bar graphs in Biber and Gray (2011) and Biber et al. (1999).

(62), premodifying nouns (40), prepositional phrases (of) (31), prepositional phrases (other) (29), and relative clauses (7). The remaining noun modifiers are all infrequent in Biber and Gray (2011) and Biber et al. (1999), whose normed frequencies are below 3.5, such as –ing clauses (2.5) and appositive NPs (3.5). It is also important to mention that the line in Figure 5 illustrates the cumulative percentages of the noun modifiers (see the orange line in the figure). Attributive adjectives, the most frequent noun modifier, comprise 42.05% of the total noun modifiers. The five frequent noun modifiers altogether comprise 93.70% of the total noun modifiers, whereas the infrequent six noun modifiers only account for 6.30%.

### 6.2. Noun Modifiers and Academic Level

Research Questions in the Set A

1. *Is there an association between academic level and NP complexity in academic writing?*
2. *What is the strength of association?*
3. *Which specific noun modifiers contribute to this association the most?*

A Chi-square test was applied to academic level and the 11 noun modifiers based on the normed frequencies of the noun modifiers. The SPSS output in Table 8 demonstrates: $\chi^2$ = 98.668, df = 10, p = 0.000. The observed $\chi^2$ value, 98.668, is greater than the critical $\chi^2$ value, 20.483 (alpha = 0.025, df = 10). Thus, there is a significant association between the 11 noun modifiers and academic level. In addition, Cramer's V is .057. However, in its focus on NPs, this dissertation adopts a fine-grained analysis only based on the 11 noun modifiers. It is not unexpected that this fine-grained analysis is associated with a small effect, but the small effect is still important. It is expected to see that 11 noun modifiers can explain 5.7% variance of the writing development from the undergraduate level to the graduate level.

Table 8. Chi-Square Test between Academic Level and NP Complexity

| | Value | DF | p value (two-sided) | Cramer's V |
|---|---|---|---|---|
| Pearson Chi-square | 98.668 | 10 | 0.000 | .057 |

*Note*. Alpha value = 0.025

A residual analysis was run as a post-hoc test to explore which specific noun modifier(s) used by students at which academic level(s) contribute to significance of the omnibus test the most. The adjusted standardized residuals were shown in Table 9. The larger the absolute values of adjusted standardized residuals are, the greater the contribution the corresponding noun modifiers contribute to the association (Sharpe, 2015). Table 9 presents that the association between the 11 noun modifiers and academic level primarily derives from attributive adjectives, premodifying nouns, relative clauses, noun complement clauses, and appositive NPs. To be specific, premodifying nouns have the largest absolute values of adjusted standardized residuals (|-8.3|, |8.3|), meaning that this noun modifier contributes to the association the most. The three other noun modifiers also have large absolute values of adjusted standardized residuals: attributive adjectives (|4.8|, |-4.8|), relative clauses (|3.2|, |-3.2|), noun complement clauses (|2.1|, |-2.1|) and appositive NPs (|-3.0|, |3.0|). For the remaining noun modifiers, the absolute values of their adjusted standardized residuals are all less than 2.00, for instance -ing clauses (|0.7|, |-0.7|) and infinitive clauses (|0.6|, |-0.6|), which suggests that these noun modifiers make little contribution to the association.

Furthermore, the adjusted standardized residuals have either positive or negative values. A positive adjusted standardized residual suggests that students at a particular academic level use a specific noun modifier more than its expected frequency, whereas a negative adjusted standardized residual indicates that students at a particular academic level use a specific noun

modifier less than its expected frequency. For the four noun modifiers above, the undergraduate students used attributive adjectives (4.8) and relative clauses (3.2), and noun complement clauses (2.1) more than their expected frequencies, whereas the graduate students used premodifying nouns (8.3) and appositive NPs (3.0) more than their expected frequencies.

Table 9. Adjusted Standardized Residuals of the Association between the Noun Modifiers and Academic Level

| | **UG** | **GR** |
|---|---|---|
| Attributive adjectives | 4.8 | -4.8 |
| Premodifying nouns | -8.3 | 8.3 |
| Relative clauses | 3.2 | -3.2 |
| -Ing clauses | -0.7 | 0.7 |
| -Ed clauses | 0.0 | 0.0 |
| Prepositional phrases (of) | 0.0 | 0.0 |
| Prepositional phrases (other) | -0.7 | 0.7 |
| Preposition + ing clauses | 0.5 | -0.5 |
| Noun complement clauses (that) | 2.1 | -2.1 |
| Infinitive clauses | 0.6 | -0.6 |
| Appositive noun phrases | -3.0 | 3.0 |

*Note*. The cells with large adjusted standardized residuals are shaded in grey.

The results reveal that NP complexity and academic level are not independent but associated with each other based on the essays in my corpus. In other words, academic level influences NP complexity. To be more specific, the influence of academic level on NP complexity derives from the influence on five noun modifiers (i.e., attributive adjectives, premodifying nouns, relative clauses, noun complement clauses, and appositive NPs) rather other noun modifiers. Although no previous studies have examined NP complexity with the same statistical perspective as in this dissertation, the findings still tend to align with an important argument that phrasal structures, primarily compressed NPs (i.e., nouns with phrasal noun modifiers), are a characteristic of graduate academic writing (Biber et al., 2011; Norris & Ortega, 2009; Wolfe-Quintero et al., 1998). For instance, Biber et al. (2011) summarize the grammatical

complexity in advanced writing as "*dense use of phrasal (nonclausal) dependent structures that function as constituents in noun phrases* (p. 29-30).

Table 10. Normed frequencies of the Four Noun Modifiers in the Undergraduate and Graduate Corpora

| | Undergraduate Corpus | Graduate Corpus | Difference |
|---|---|---|---|
| Attributive adjectives* | 6235 | 6532 | 297 |
| Premodifying nouns* | 1618 | 2326 | 708 |
| Relative clauses* | 781 | 744 | -37 |
| Noun complement clauses (that) | 134 | 115 | -19 |
| Appositive noun phrases | 108 | 174 | 66 |

Note. The normed frequencies represent the frequencies of the noun modifiers in 100,000 words. The asterisks mark the frequent noun modifiers in the main corpus. The negative values mean that the graduate corpus has fewer of a noun modifier than the undergraduate corpus.

The residual analysis only pinpoints the source of influence of academic levels on NP complexity, but it does not reveal the comparison of the four modifiers between the undergraduate and graduate levels. Table 10 demonstrates the observed (normed) frequencies of the four noun modifiers in the undergraduate corpus and the graduate corpus. There are 297 attributive adjectives, 708 premodifying nouns, and 66 appositive NPs in the graduate corpus more than in the undergraduate corpus. In contrast, there are 37 relative clauses and 19 noun complement clauses in the undergraduate corpus more than in the graduate corpus. The differences on the observed (normed) frequencies of the five noun modifiers suggests how academic level influences NP complexity in my corpus: graduate students tend to use more attributive adjectives, premodifying nouns, and appositive NPs to build compressed NPs, whereas undergraduate students use more relative clauses and noun complement clauses. This pattern has been argued indicative of writing development: graduate students use significantly more phrasal noun modifiers (e.g., premodifying nouns and appositive NPs) in their academic

writing, whereas undergraduate students use significantly more relative clauses and noun complement clauses (Parkinson & Musgrave, 2014; Staples et al. 2016).

Considering the adjusted standardized residuals and the observed (normed) frequencies of the noun modifiers in my corpus, I found that academic level primarily influences NP construction due to the influence on attributive adjectives, premodifying nouns, relative clauses, noun complement clauses, and appositive NPs. Having said this, the use of the five noun modifiers should be analyzed from a qualitative perspective, because writing development on grammatical complexity between the two academic levels cannot be simply concluded based on greater and/or fewer uses of these noun modifiers.

Excerpt 1 and Excerpt 2 were from the randomly selected essays in the undergraduate corpus and the graduate corpus, to illustrate the NP construction in the two corpora, respectively (see the excerpts below). In comparison, Excerpt 1 includes more clausal structures than Except 2, such as coordinate clauses, verb complement clauses, adverbial clauses, and a clausal noun modifier (i.e., relative clauses). Relative clauses in Excerpt 1 includes a restrictive relative clause (*who are willing to take part in the political process*) and a nonrestrictive relative clause (*each of which has a distinct impact on political consistency*). The remaining noun modifiers are primarily attributive adjectives (e.g., *ideal*, *political*, and *wider*). In contrast, Excerpt 2 are highly based on compressed NPs, and clausal structures are absent. For instance, the first sentence is *Financial Services Compensation Scheme (FSCS) of the Financial Service and Markets Act 2000 (Collective Investment Schemes) Order 2001 (SI2001/1062) is the UK official fund of last resort for customers of authorized persons.* This sentence has only one verb (*is*), and both the subject and object are compressed NPs, with attributive adjectives (e.g., *financial*, *collective*, *authorized*), premodifying nouns (e.g., *compensation*, *investment*, *Markets*), and appositive NPs

(e.g., *FSCS, SI2001/1062*). Compressed NPs are more frequent in the graduate essays (i.e., the advanced-level writing) in the corpus.

**Excerpt 1** (L2 undergraduate essay, file no. 0138a)

- The investigators identify three <ideal> types *of* <political> culture, each of which has a distinct impact on <political> consistency. In a <parochial> culture, the electorate is ignorant of <political> events and rarely becomes involved in <political> activity; in a [subject] culture, citizens have <wider> knowledge, but feel politically impotent; and a [participant] culture comprises highly informed actors who are willing to take part in the <political> process. The authors explain that a [participant] culture can also weaken the <political> system by causing a '<democratic> overload', where <excessive> pluralism hinders the <decision-making> process. This conclusion establishes a <direct> <causal> chain between <political> culture and the permanence *of* <democratic> systems.

*Note*. Attributive adjectives are in angled brackets, premodifying nouns in square brackets, relative clauses are underlined, and prepositional phrases are in italics.

**Excerpt 2** (L2 graduate essay, file no. 0428b)

- <Financial> [Services] [Compensation] Scheme **(FSCS)** *of* the <Financial> [Services] and [Markets] Act 2000 **(<Collective> [Investment] Schemes)** Order 2001 **(SI2001/1062)** is the [UK] <official> fund *of* <last> resort *for* customers *of* <authorised> persons. In principle, it applies only to <UK-based> [retail] institutions: therefore, <overseas-controlled> funds aimed at <sophisticated> investors, such as [hedge] funds, fall outside its scope. In any event, the threshold *for* compensation **(£48,000)** would not be of much help. The rules *in* the [FSA's] Conduct *of* [Business] [sourcebook] cover, inter alia, [business] promotion, [business] policy, advise standards, dealings.

*Note*. Attributive adjectives are in angled brackets, premodifying nouns in square brackets, appositive NPs in bold, and prepositional phrases in italics.

In addition, the excerpts also demonstrate that topic has a large influence on the two phrasal noun modifiers (i.e., attributive adjectives and premodifying nouns), which are often related compressed NPs in advanced writing, for instance the graduate essays. Biber and Conrad (2009) identified *topic* as one of the six main situational characteristics that can influence language use. However, in corpus studies, it is hard to eliminate the topic influence because it is hard to build a large-scale corpus with writing on the same topic. Staples and Reppen (2016) demonstrate that the same cases of attributive adjectives and premodifying nouns regarding certain content (e.g., video games, social network) are repeatedly used in their corpus of

academic writing, and they point out that "we should be mindful that greater use does not always mean greater development" (p. 30). The repetition of attributive adjectives and premodifying nouns can also be seen in the two excerpts. Excerpt 1 presents that *political* was used multiple times to modify different nouns, as in *political culture*, *political events*, *political,* and *political system*. Similarly, Excerpt 2 demonstrates that *business* was used multiple times to modify different nouns, as in *business sourcebook*, *business promotion* and *business policy*. The two instances of phrasal noun modifiers (*political*, *business*) are repeatedly used to help the content development throughout the two excerpts (and the two selected essays). For instance, Excerpt 1 summarizes three types of ideal political culture. For each type of political culture, the author discusses political events, political activities and political process. As a result, the excerpt demonstrates that the content of political culture is developed based on the use of this attributive adjective (*political*).

Overall, from the frequency perspective, the influence of academic writing on NP complexity is primarily based on the influence on attributive adjectives, premodifying nouns, relative clauses, noun complement clauses, and appositive NPs in my corpus. Graduate students use more of the three phrasal modifiers and fewer of the two clausal modifiers in my corpus. In opposite, undergraduate students use more of the two clausal modifiers and fewer of the three phrasal modifiers in my corpus. This adds empirical support to the claim that compressed NPs are a characteristic of advanced academic writing. However, the repetition of using content-related attributive adjectives and premodifying nouns might influence this sequence of writing development, which is worth being investigated further.

### 6.3. Noun Modifiers and L1- and L2-English Language Background

Research Questions in the Set B

1. *Is there an association between L1- and L2-English language background and NP complexity in academic writing?*
2. *What is the strength of association?*
3. *Which specific noun modifiers contribute to this association the most?*

A Chi-square test was run to investigate L1- and L2-English language background and the 11 noun modifiers based on the normed frequencies of the noun modifiers. The SPSS output is shown in Table 11: $\chi^2 = 115.559$, df = 10, p = 0.000. The observed $\chi^2$ value, 115.559, is greater than the critical $\chi^2$ value, 20.483 (alpha = 0.025, df = 10). Thus, there is a significant association between L1- and L2-English language background and the use of noun modifiers. In addition, Cramer's V is .062. However, in its focus on NPs, this dissertation adopts a fine-grained analysis only based on the 11 noun modifiers. Again, it is expected that a fine-grained analysis is associated with a small effect, but the small effect is not unimportant. It is expected to see only the use of the 11 noun modifiers can explain 6.2 % variance of writing difference between L1 and L2 students.

Table 11. Chi-Square Test between L1- and L2-English Language Background and NP Complexity

| | Value | DF | p value (two-sided) | Cramer's V |
|---|---|---|---|---|
| Pearson Chi-square | 115.559 | 10 | .000 | .062 |

*Note.* Alpha value = 0.025

Next, the residual analysis was run as a post-hoc test to explore which specific noun modifier(s) used by students from which L1- and L2-English language background contribute the

most to the significance of the omnibus test the most. The standardized residuals were reported in Table 12. In contrast to the association with academic level (see Table 9), eight noun modifiers largely contribute to the association, whereas only three noun modifiers (i.e., -ing clauses, -ed clauses, and prepositional phrases [of]) make little contribute to the association. Table 12 presents that premodifying nouns have the largest absolute values of adjusted standardized residuals (|-6.5|, |6.5|), suggesting that this noun modifier contributes to the association the most. Seven other noun modifiers also have large absolute values of adjusted standardized residuals: attributive adjectives (|-2.3|, |2.3|), relative clauses (|3.0|, |-3.0|), prepositional phrases (other) (|5.8|, |-5.8|), prepositional phrases followed by -ing clauses (|2.1|, |-2.1|), noun complement clauses (|2.6|, |-2.6|), infinitive clauses (|2.6|, |-2.6|) and appositive NPs (|-3.8|, |3.8|). This means that these seven noun modifiers also make large contributions to the association. Moreover, for the eight noun modifiers above, the L1 students used five noun modifiers more than their expected frequencies: relative clauses (3.0), prepositional phrases (other) (5.8), prepositional phrases followed by -ing clauses (2.1), noun complement clauses (2.6) and infinitive clauses (2.6). The L2 students only used three noun modifiers more than their expected frequencies, namely, attributive adjectives (2.3), premodifying nouns (6.5) and appositive NPs (3.8).

Table 12. Adjusted Standardized Residuals of the Association between the 11 Noun Modifiers and L1- and L2-English Language Background

| | L1 | L2 |
|---|---|---|
| Attributive adjectives | -2.3 | 2.3 |
| Premodifying nouns | -6.5 | 6.5 |
| Relative clauses | 3.0 | -3.0 |
| -Ing clauses | 0.4 | -0.4 |
| -Ed clauses | -1.8 | 1.8 |
| Prepositional phrases (of) | 1.7 | -1.7 |
| Prepositional phrases (other) | 5.8 | -5.8 |
| Preposition + ing clauses | 2.1 | -2.1 |
| Noun complement clauses (that) | 2.6 | -2.6 |
| Infinitive clauses | 2.6 | -2.6 |
| Appositive noun phrases | -3.8 | 3.8 |

*Note*. The cells with large standardized residuals are shared in grey in the table.

The results reveal that NP complexity and L1- and L2-English language background are not independent but associated with each other. Not surprisingly, L1- and L2-English language background influences NP complexity in my corpus, because native languages are often transferred into L2 acquisition in either a positive or negative way (Hall, 2011). In terms of the source of this influence, a noteworthy point is: the influence of L1- and L2-English language background on NP complexity in my corpus derives from a much wider range of noun modifiers than the influence of academic level, including the eight aforementioned noun modifiers, for instance attributive adjectives, relative clauses, prepositional phrases (other), and infinitive clauses. This wide range of noun modifiers suggests that compared to academic level, L1- and L2-English language background tends to influence the diversity of NP patterns instead of NP compression in my corpus.

The residual analysis can identify the source of the influence of L1- and L2-English language background on NP complexity, but it does not show the comparison of the eight modifiers between the L1 and L2 backgrounds. Table 13 demonstrates the observed (normed) frequencies of these eight noun modifiers in the L1 and L2 corpora. There are 69 relative clauses,

220 prepositional phrases (other), 29 prepositional phrases followed by –ing clauses, 34 noun complement clauses, 36 infinitive clauses in the L1 corpus more than in the L2 corpus. In contrast, there are 566 attributive adjectives, 509 premodifying nouns and 73 appositive NPs in the L2 corpus more than in the L1 corpus. The differences on the observed (normed) frequencies of the eight noun modifiers suggests how L1- and L2-English language background influences NP complexity in my corpus: L1 students tend to use noun modifiers with a broader range to construct more diverse NP patterns than L2 students do. However, it is surprising to see that L2 students use more attributive adjectives, premodifying nouns and appositive NPs, which are often used to build compressed NPs, a characteristic of advanced writing.

Deng, Lei and Liu (2020) have recently published a synthesized article in *Applied Linguistics* to comment on the development of grammatical complexity between L1 and L2 students in academic writing.

> English as a Second or Foreign Language (ESL/EFL) graduate students' writing displayed a higher level of syntactic sophistication than their English L1 counterparts' writing, which might be difficult to fathom given that L1 writers generally possess a higher language proficiency than L2 writers at the same educational level[28] (p. 2).

The term *syntactic sophistication* refers to the grammatical complexity displayed in advanced academic writing (e.g., graduate essays). It is not the same as *grammatical sophistication* defined as the depth of grammatical construction (see Chapter 2 of this dissertation) but is similar to *structural compression* in Biber and Gray (2010) and/or *phrasal complexity* in Biber et al. (2011). From the perspective of frequency analysis, the quantitative finding aligns with Deng, Lei and Liu's (2020) claim, because more of the three phrasal modifiers (i.e., attributive adjectives, premodifying nouns and appositive NPs) are used in the L2 corpus than in the L1

[28] Deng, Lei and Liu (2020)'s claim is based on Lei and Liu (2015), an unpublished writing project presented at the American Association for Applied Linguistics Conference in 2015.

corpus. However, the use of the eight noun modifiers should be explored from a qualitative perspective and writing difference on grammatical complexity between L1 and L2 students cannot be simply concluded based on greater and/or fewer uses of these noun modifiers.

Table 13. Normed Frequencies of the Four Noun Modifiers in the L1 and L2 Corpora

| | L1 Corpus | L2 Corpus | Difference |
|---|---|---|---|
| Attributive adjectives* | 6131 | 6697 | -566 |
| Premodifying nouns* | 1802 | 2311 | -509 |
| Relative clauses* | 793 | 724 | 69 |
| Prepositional phrases (other)* | 1840 | 1620 | 220 |
| Preposition + ing clauses | 192 | 163 | 29 |
| Noun complement clauses (that) | 139 | 105 | 34 |
| Infinitive clauses | 172 | 136 | 36 |
| Appositive noun phrases | 112 | 185 | -73 |

Note. The normed frequencies represent the frequencies of the noun modifiers in 100,000 words. The asterisks mark the frequent noun modifiers in the main corpus. The negative values show the noun modifiers that are used more in the L2 corpus than in the L1 corpus.

Excerpt 3 and Excerpt 4 were randomly selected to illustrate the NP construction in the L1 corpus (at the undergraduate and graduate levels). The two excerpts can be compared with Excerpt 1 and Excerpt 2 which demonstrate NP construction in the L2 corpus (at the undergraduate and graduate levels) in Section 6.2. In comparison, the L1 corpus has more diverse noun modifiers than the L2 corpus. To be more specific, Excerpt 3 (from an L1 undergraduate essay) includes five types of noun modifiers that occur in undergraduate essays: attributive adjectives (e.g., *impartial*), premodifying nouns (e.g., *historian's*), prepositional phrases (e.g., *the social functions of these memorials*), relative clauses (e.g., *that has created them*, and noun complement clauses, for example *the assumption that the Holocaust should be memorialised*. In contrast, Excerpt 1 (from an L2 undergraduate essay) includes only three noun modifiers, namely attributive adjectives, premodifying nouns, and appositive NPs.

**Text Except 3** (L1 undergraduate students, file no. 0019a)

- The role *of* the <impartial> judge is one that is hard to play, and though it is acknowledged that the [historian's] task is not to judge, it is not entirely possible for the historian to distance herself from this issue, because she is necessarily the product *of* her time. As such, the writer attempts to bring in the professionalism *of* her discipline to tackle this issue by first calling into question the assumption that the Holocaust should be memorialised; the <social> functions *of* these memorials; and finally, by examining the <various> memorials, aim to show that the forms that memorials take often reflect their position in time and location, and more importantly, is a reflection *of* the society that has created them, because the form that a memorial takes is not dictated by a <single> <universal> principle, but obeys a number *of* <intertwined> forces.

*Note*. attributive adjectives are in angle brackets, premodifying nouns in square brackets, relative clauses and noun complement clauses are underlined, and prepositional phrases are in italics.

This is also the case for the L1 and L2 graduate essays. Similarly, Excerpt 4 (from a L1 graduate essay) shows six types of noun modifiers: attributive adjectives (e.g., *disproportionate*), premodifying nouns (e.g., *phoneme level*), prepositional phrases (e.g., *his emphasis on the chart*), prepositional phrases followed by -ing clauses (i.e., *the only way of introducing pronunciation*), -ing clauses (i.e., *time being spent on teaching at the phoneme level*), -ed clauses (i.e., *the sounds commonly used in English*), and relative clauses (e.g., *in which his phonemic chart can be used*). In contrast, Excerpt 2 (from an L2 graduate essay) only includes four main noun modifiers, which are attributive adjectives, premdifying nouns, appositive NPs, and prepositional phases (other).

**Text Excerpt 4** (L1 graduate students, file no. 3118 a)

- Underhill suggests ways in which his <phonemic> chart can be used at <different> stages *of* the lesson, with <training> and <awareness-raising> activities at the beginning *of* course. As someone who has used this chart for many years *of* teaching, I feel that this is a <useful> tool in training both teachers and students in the sounds commonly used in English. I do not however consider that it should be the only way *of introducing* pronunciation into the <Academic> [English] syllabus. Although Underhill does devote part *of* the book to the <supra-segmental> features *of* intonation, stress and other features *of* <connected> speech, it might be true that his emphasis *on* the chart encourages a <disproportionate> time being spent on teaching at the [phoneme] level.

*Note*. attributive adjectives are in angle brackets, premodifying nouns in square brackets, relative clauses and -ing/-ed clauses are underlined, and prepositional phrases (followed by -ing clauses) are in italics.

The qualitative analysis suggests that in my corpus, L1 students tend to use broader range of noun modifiers than L2 students at both the undergraduate level and graduate level. Although no previous studies directly compared L1 and L2 writing on these noun modifiers, the findings align with Staples and Reppen (2016), who argue that compared to L1 English students, L2 students rely on repetition use of attributive adjectives and premodifying nouns. The argument in Staples and Reppen (2016) can also be supported by the repeated attributive adjectives (*political*) and premodifying nouns (*business*) in Excerpt 1 and Excerpt 2. As a result, the claim in Deng, Lei and Liu (2020) that L2 students write in a more sophisticated manner based on compressed NPs needs a further investigation, because the repetition of phrasal noun modifiers (i.e., attributive adjectives, premodifying nouns, appositive NPs) may not suggest a sophisticated manner of writing.

### 6.4. Clarification on Prepositional Phrases (of)

Prepositional phrases (of) is worth being discussed to clarify a potential misunderstanding. Based on *Chi-square* and *residual analysis*, prepositional phrases (of) do not contribute to the influence of academic level and the influence of the L1- and L2-English language background on NP complexity, respectively. However, this does not mean that prepositional phrases (of) are not used to construct NPs in my corpus. It is the second most frequent noun modifier in my corpus (see Figure 5) and all the four excerpts include multiple prepositional phrases (of) in the NPs, two cases in Excerpt 1, four cases in Excerpt 2, six cases in Excerpt 4, and six cases in Excerpt 4. The examples are:

- *the permanence of democratic systems* (in Excerpt 1)
- *official fund of last resort* (in Excerpt 2)

- *the role of the impartial judge* (in Excerpt 3)
- *the supra-segmental features of intonation* (in Excerpt 4)

The small adjusted standardized residuals and high frequencies of the prepositional phrases suggest that prepositional phrases (of) is used as a frequent noun modifier to construct NPs, because its frequency is *not unexpected*. The adjusted standardized residuals are small, that is, expected use aligns with actual frequency of use in the corpus. A tentative interpretation is that prepositional phrases (of), as a noun modifier, have a wide range of functions in English, which makes this noun modifier naturally pervasive in academic writing. As noted by Biber et al. (1999), "academic prose is noteworthy, for a large number of noun + of-phrase expressions are used repeatedly" (p. 296). This is due to "the extremely wide range of functions for this preposition" (p. 295), including but not limited to quantity description (e.g., loads of), species description (e.g., type of), physical description (e.g., the position of), existence (e.g., the presence of), abstract qualities (e.g., the value of) and process (e.g., the developmental of). Due to these diverse functions, it is not surprising that this noun modifier is frequently used in academic writing in my corpus. Therefore, although preposition phrases (of) have little contribution to the influence of both academic level and L1- and L2-English language background on NP complexity, it is still frequently used to construct NPs in my corpus.

# CHAPTER 7. CONCLUSIONS AND IMPLICATIONS

In this chapter, I summarize the primary research findings of this dissertation by answering the research questions. Then, I provide research implications on grammatical complexity and pedagogical implications for instruction on NP complexity in writing classroom. Last, limitations of this dissertation are discussed, and possible solutions are offered.

## 7.1. Summary of Research Findings

The purpose of this dissertation is to study how NP complexity is influenced by academic level and L1- and L2-English language background in academic writing. Based on the register perspective, NP complexity is operationalized to the 11 noun modifiers in Biber et al.'s (2011) hypothesized developmental index of writing complexity features. This fine-grained investigation allows me to investigate the unique contributions of the individual noun modifiers to NP complexity in my corpus. The first major finding is that academic level influences NP complexity, and this influence derives from the influence on attributive adjectives, premodifying nouns, relative clauses, noun complement clauses, and appositive NPs in my corpus. Based on the essays in my corpus, graduate students tend to use more phrasal modifiers (i.e., attributive adjectives, premodifying nouns, appositive NPs), whereas undergraduate students produce more clausal modifiers (i.e., relative clauses and noun complement clauses). This aligns with the important previous argument that compressed NPs are a characteristic of advanced academic writing (Biber et al., 2011; Norris & Ortega, 2009; Wolfe-Quintero et al., 1998). The second major finding is that L1- and L2-English language background influences NP complexity. In contrast to academic level, this influence of L1- and L2-English language background derives from a broader range of noun modifiers, which includes attributive adjectives, premodifying

nouns, relative clauses, prepositional phrases (other), prepositional phrases followed by -ing clauses, noun complement clauses, infinitive clauses, and appositive NPs. Based on the essays in my corpus, L1 students tend to build more diverse NP patterns with various noun modifiers than L2 students. Surprisingly, L2 students use more attributive adjectives, premodifying nouns, and appositive NPs, the grammatical features of advanced academic writing. A qualitative analysis demonstrates that L2 students repeatedly use the same cases of the phrasal noun modifiers for content development in their essays. I agree with the claim that greater use does not always mean greater development and/or difference (Staples & Reppen, 2016). More research attention should be paid to these repetitions in future's research. As this dissertation is based on a nonparametric test, I plan to replicate it with different writing samples in the future to see whether the two major findings are reliable or not.

### 7.2. Research Implications

The dissertation is descriptive and exploratory; nevertheless, it adds an additional piece of evidence to the importance of NPs in academic writing. Via a corpus-based approach, this dissertation demonstrates the different types of influence that academic level and L1- and L2-English language background have on NP complexity in academic writing. In writing research, there are opposing opinions on whether NPs should be generally taught in academic writing courses. For instance, some argue that compressed NPs (i.e., nominal structures) are likely to cause learning difficulties in academic contexts, so this structure should receive instruction in EAP courses (Fang, Schleppegrell, & Cox, 2006; Musgrave & Parkinson, 2014). In contrast, Tucker-McLaughlin and Hubbard (2013) suggest using the "paramedic" method: for the sake of clarity, students should avoid using compressed NPs by reducing phrasal noun modifiers (e.g., prepositional phrases). Because of these differing opinions among writing researchers, it is

critical to explore NP complexity further in order to clarify the necessity and importance of teaching NPs.

Based on these major findings of this dissertation, a reasonable assumption is that the interaction of academic level and L1- and L2-English language background may influence NP complexity in academic writing. This interaction includes four specific levels, namely L1-undergraduate, L2-undergradaute, L1-gradaute, and L2-gradaute. It is interesting to explore the trajectory of writing development on grammatical complexity based on the interaction. Table 14 demonstrates the developmental differences on the normed frequencies of the 11 noun modifiers between the undergraduate and graduate levels of L1 and L2 English language students, respectively. For instance, the L2 developmental difference was calculated by dividing the normed frequencies of the 11 noun modifiers in the L2 undergraduate corpus from the L2 graduate corpus. The asterisks mark the five noun modifiers that follows an opposite developmental pattern. The L1 graduate corpus has 903.40 attributive adjectives more than the L1 undergraduate corpus, whereas the L2 graduate corpus has then 265.58 attributive adjectives less than the L2 undergraduate corpus. This opposite developmental pattern can also be seen from relative clauses, prepositional phrases (other), prepositional phrases followed by -ing clauses, and infinitive clauses.

Table 14. Developmental Differences of the Noun Modifiers between L1 and L2 Students

| Noun Modifiers | L1 Developmental Difference | L2 Developmental Difference |
|---|---|---|
| Attributive adjectives* | 903.40 | -265.58 |
| Premodifying nouns | 539.33 | 871.02 |
| Relative clauses* | -97.98 | 18.75 |
| -Ing clauses | 45.30 | 11.39 |
| -Ed clauses | 11.67 | 35.79 |
| Prepositional phrases (of) | 352.93 | 384.83 |
| Prepositional phrases (other)* | -212.08 | 635.33 |
| Preposition + ing clauses* | -38.17 | 56.08 |
| Noun complement clauses (that) | -31.41 | -8.10 |
| Infinitive clauses* | -20.96 | 32.24 |
| Appositive noun phrases | 45.60 | 84.53 |

Note. The normed frequencies represent the frequencies of the noun modifiers in 100,000 words.

From the undergraduate level to the graduate level, the writing development on grammatical complexity might not be the same between L1 and L2 students. So far, I have two hypotheses that L1 and L2 students might (1) either follow different trajectories of writing development on grammatical complexity or (2) follow the same trajectory of writing development on grammatical complexity but with different paces. For instance, the writing development of L1 students could be faster than L2 students, or vice versa. So far, few studies have focus on clarifying the trajectory of writing development on grammatical complexity based on the interaction between academic level and L1- and L2-English language background. I plan to conduct follow-up studies to examine my two hypotheses in the future.

### 7.3. Limitations and Possible Solutions

The dissertation is not without limitations. First, the corpus in the dissertation includes a single written genre (i.e., essays). Although this is a common written genre, it cannot comprehensively represent academic writing in universities. Other genres can be included in future research, such as proposals, argumentative papers, and case studies, among others. Second, there is a topic influence on the use of the noun modifiers, especially attributive

adjectives and premodifying nouns. Test essays are often based on the same topic(s), which can effectively eliminate the topic influence. I plan to write research proposals to either testing companies (e.g., ETS) or universities that have testing programs to see if I can use test essays in my future's studies. Third, the core value of residual analysis needs to be further explored. In the methods, I present the calculation process of residual analysis and mention the new angle of frequency analysis that this statistical method can add. However, it is important to compare residual analysis with the other alternative methods that are used as post-hoc tests in statistics. This comparison would help me better justify the use of residual analysis in similar research. Fourth, in terms of the noun modifiers, I only studied relative clauses with relativizers. Syntactic parsing (constituency parsing) is a promising way of identifying relative clauses with zero relativizers in future research. It is also interesting to explore how to build a program to extract noun modifiers based on both part-of-speech and syntactic parsed information, to deal with the noun modifiers whose grammatical functions cannot be tagged by the *Biber Tagger*. Fifth, this dissertation is based on a nonparametric test (*Chi-square test*) because of the non-normal distributions of some noun modifiers, so the findings are sample-driven and cannot be generalized to other research contexts. Also, this nonparametric test cannot capture the variability among the writers in this study. However, it does reveal the different uses of NPs among different groups of writers (i.e., graduate and undergraduate students, L1 and L2 students). For research based on nonparametric tests, replication would be an important way to explore whether the major findings in my dissertation tend to be reliable. When I replicate my dissertation in the future, I also plan to enhance the research design, by which I can test my two aforementioned hypotheses on the trajectory of writing development on grammatical complexity between L1 and L2 students.

# APPENDIX A. SAMPLE TAGGED FILE

The ^ati++++=The
development ^nn+nom+++=development
of ^in++++=of
feminist ^nn+nom+++=feminist
thought ^vbd+++xvbn+=thought
and ^cc++++=and
action ^nn++++=action
in ^in++++=in
Japan ^np++++=Japan
is ^vbz+bez+vrb++=is
distinct ^jj+pred+++=distinct
from ^in++++=from
that ^dt+dem+++=that
of ^in++++=of
its ^pp$+it+++=its
emergence ^nn+nom+++=emergence
in ^in++++=in
the ^ati++++=the
West ^np+pl+++=West,
, ^zz++++=EXTRAWORD
more ^rbr++++=more
specifically ^rb++++=specifically
within ^in++++=within
the ^ati++++=the
United ^np++++=United
States ^nns++++=States
and ^cc++++=and
Britain ^np++++=Britain.
. ^zz++++=EXTRAWORD

# APPENDIX B. CALCULATION OF PRECSION AND RECALL

The tags of the target features are put into four specific categories and denoted by "0" and "1":

1. annotated - not modified → [1, 0]
2. annotated - modified → [1, 1]
3. not annotated - not modified → [0, 0]
4. not annotated - modified → [0, 1]

The table below provides examples for each of the four categories.

1. *Necessary* [1, 0]: the gold label is consistent with the tag, and the tag does not need to be modified.
2. *Highly* [1, 1]: the gold label is not consistent with the tag, and the tag needs to be modified (from an attributive adjective to an adverb).
3. *Is* [0, 0]: the gold label is consistent with the tag. *Is* is not tagged as a target feature, and it does not need to be modified.
4. *Exchange* [0, 1]: the gold label is not consistent with the tag. *Exchange* is not tagged as a target feature, and the tag needs to be modified (from a verb to a noun)

*Sample Tags and Tag Checking*

| File | Word | Gold Label | Tags | Annotated | Modified |
|---|---|---|---|---|---|
| 0050a | highly | | jj+atrb+++=highly | 1 | 1 |
| | necessary | ADJ | jj+atrb+++=necessary | 1 | 0 |
| | condition | N | nn+nom+++=condition | 1 | 0 |
| | for | PREP | in++++=for | 1 | 0 |
| | exchange | N | vb++++=exchange | 0 | 1 |
| | is | | vbz+bez+vrb++=is | 0 | 0 |

The four categories of the tags can be converted to the four relationships between gold labels and tags mentioned in Pustejovsky and Stubbs (2012):

1. true positive → [1, 0]
2. false positive → [1, 1]
3. true negative → [0, 0]
4. false negative → [0, 1]

The formula in Pustejovsky and Stubbs (2012) to calculate precision and recall can be converted to the formula below, which are used to calculate the precision and recall for all the target features.

- Precision = true positive / (true positive +false positive)=sum[1,0]/(sum[1,0]+sum[1,1])
- Recall = true positive / true positive +false negative)=sum[1,0]/(sum[1,0]+sum[0,1])

# APPENDIX C. TRAINING MATERIALS OF TAG CHECKING

Instructions on checking the five chunking patterns:

- **Noun + noun sequence**
    - **Two-noun sequences**
        - Check if a noun-noun sequence can be converted to "of genitive"
            - Computer science → "science of computer"
            - If yes, keep it. If not, delete it.
    - **Multiple-noun sequences**
        - Check if a noun-noun sequence can be converted to "of genitive" for each pair of a two-noun sequence
            - Corpus research methods → "methods of corpus research"
            - Corpus research methods → "research methods of corpus"
            - If yes, keep them. If not, delete them.
    - **Coordinate structures**
        - Check if a coordinate structure can be converted to "of genitive"
            - Speech and pause rates → "rate of speech and pause"
            - If yes, keep it. If not, delete it
- **Noun + preposition (and Noun + preposition + ing)**
    - Noun + preposition sequence
        - Check if a noun + preposition sequence can be converted to a noun + relative clause (without changing the meaning).

- The book on the table is not mine → "the book which is one the table is not mine"?
- I put my book on the table → "I put my book which is one the table".
- If yes, keep it. If not, delete it.

▪ Check if the preposition is part of an NP or part of a VP.

- The book on the table is not mine → "[the book on the table]$^{NP}$ is not mine"
- I put my book on the table → I [put my book on the table]$^{VP}$
- If NP, keep it. If VP, delete it.

o Noun + preposition + ing

▪ Refer to the instruction for noun + preposition sequence above.

- **Noun + infinitive clause (to)**

o Noun + infinitive clause (to)

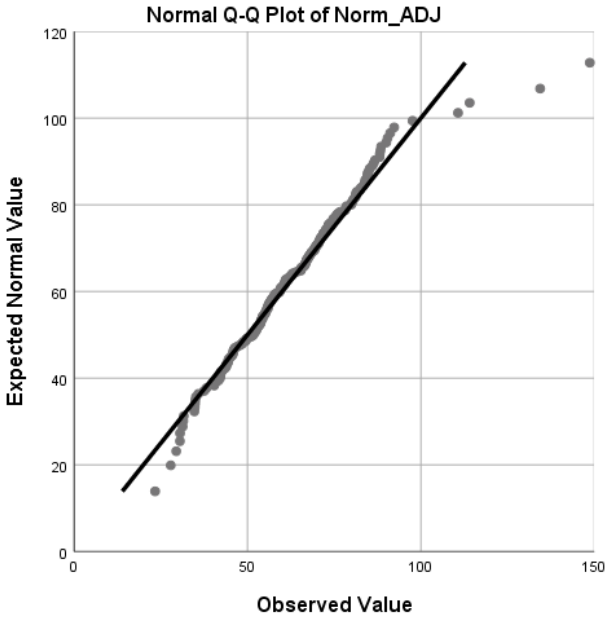
▪ Check if a noun + infinitive clause sequence can be converted to a noun + relative clause (without changing the meaning).

▪ The method to detect the problem is wrong → "the way which is used to detect the problem is wrong"

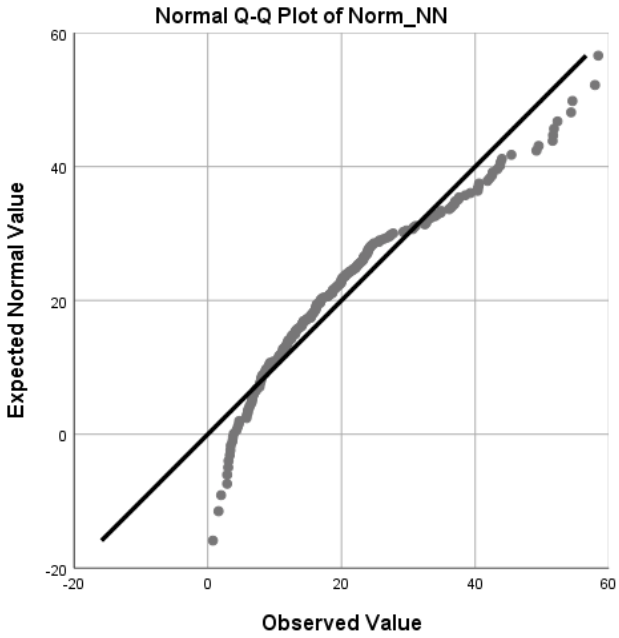
- If yes, keep it. If not, delete it.

# APPENDIX D. QQ PLOTS OF THE 11 NOUN MODIFIERS

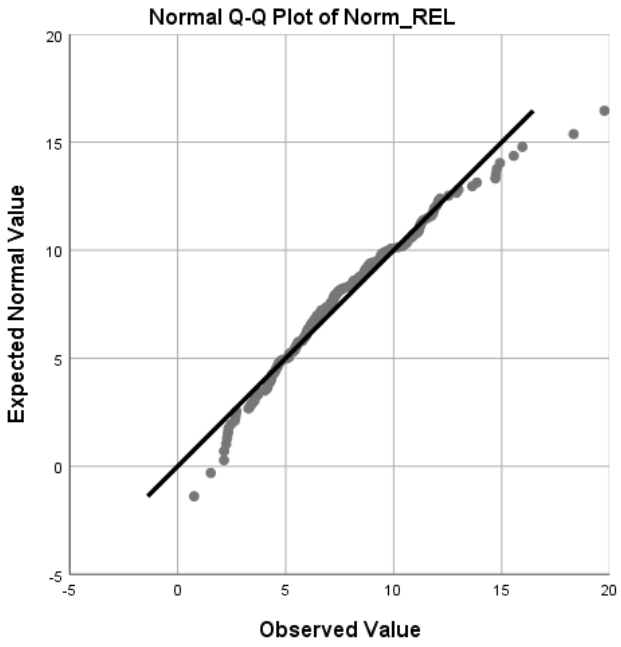
Attributive adjectives



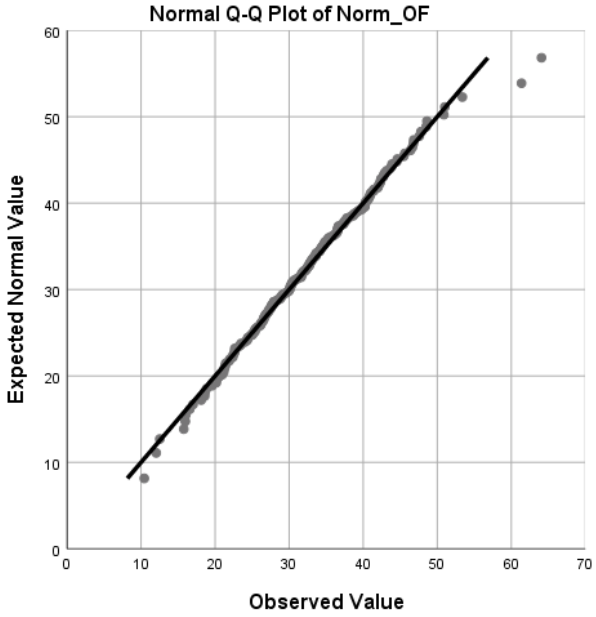

Premodifying nouns



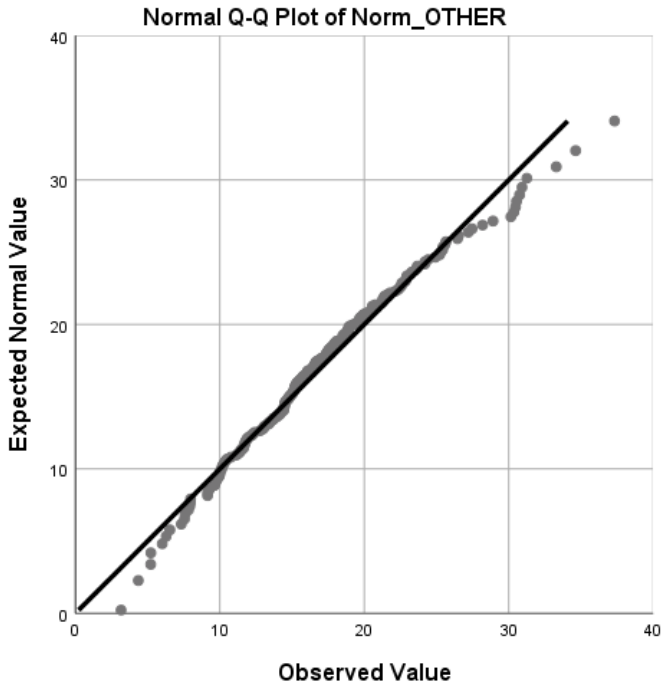
Relative clauses



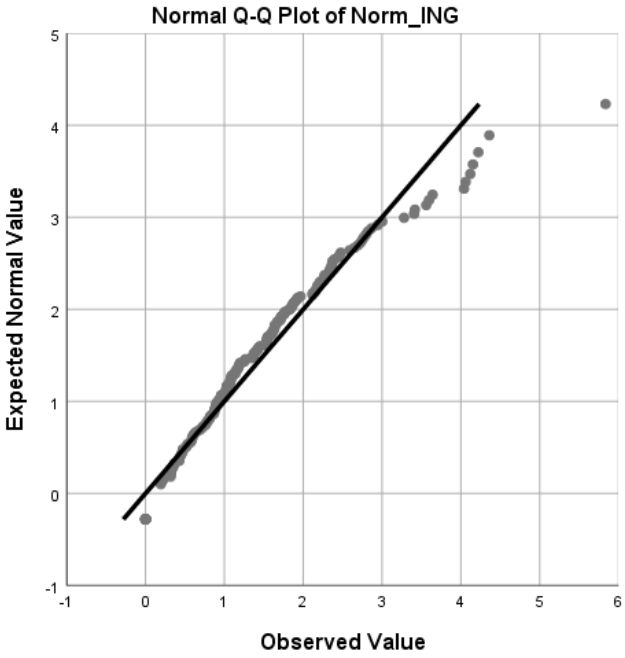

Prepositional phrases (of)



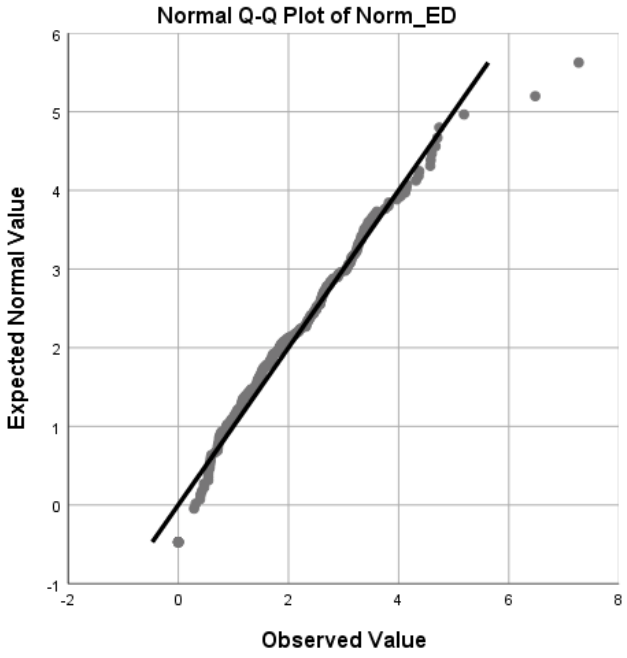
Prepositional phrases (other)



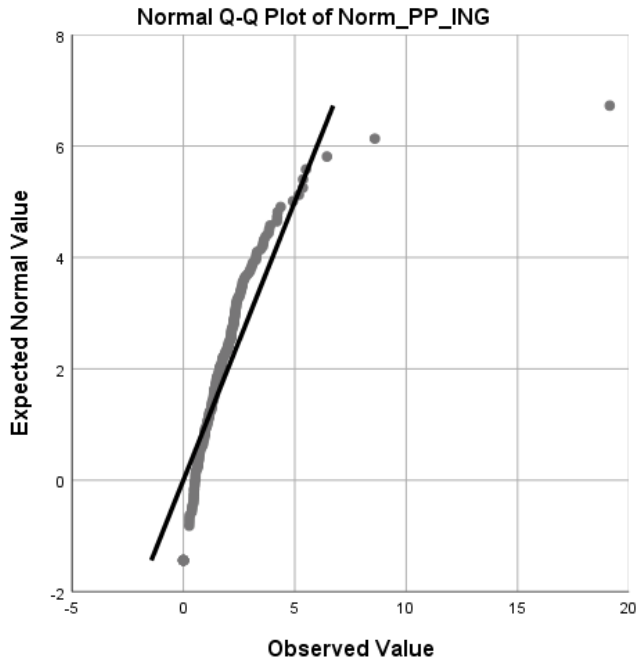

-Ing clauses as noun modifiers



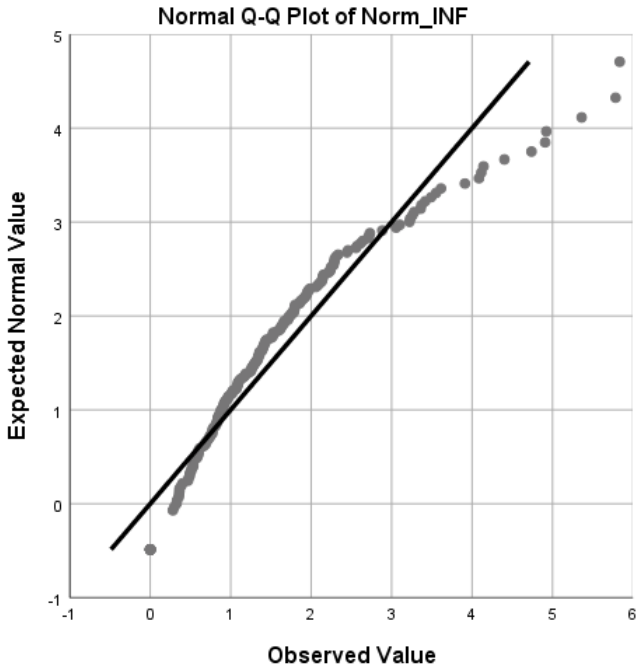
-Ed clauses as noun modifiers



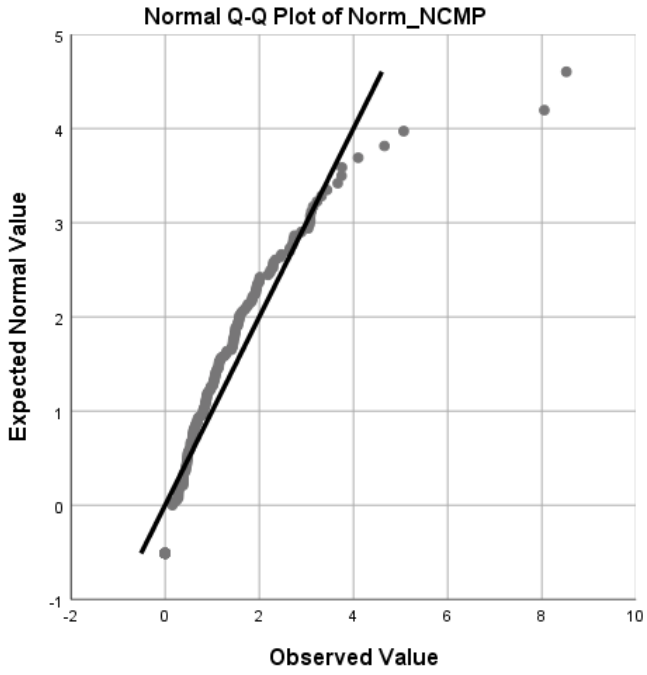

Prepositional phrases followed by –ing clauses



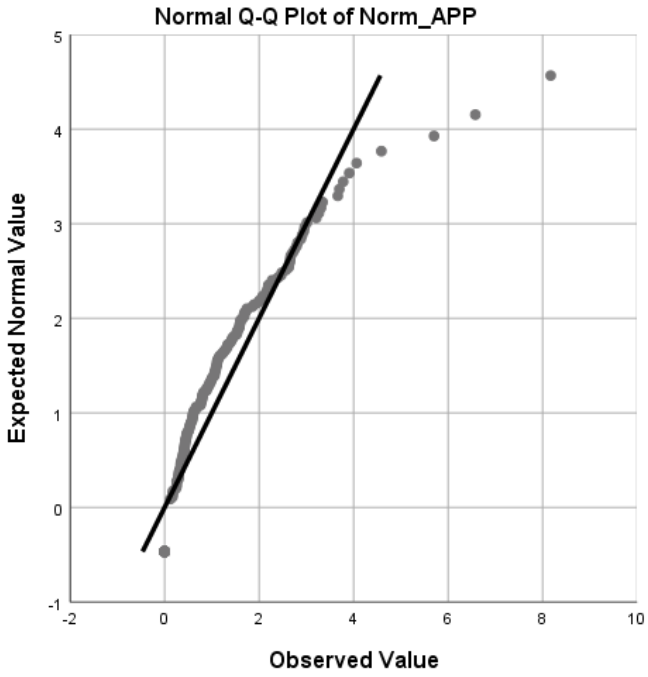
Infinitive clauses as noun modifiers




Noun complement clauses



Appositive noun phrases



# REFERENCES


Bardovi-Harlig, K. (1992). A second look at T-unit analysis: reconsidering the sentence. *TESOL Quarterly*, *26*, 390–395.

Berlage, E. (2014). *Noun phrase complexity in English*. Cambridge, United Kingdom: Cambridge University Press.

Biber, D. (1986). Spoken and written textual dimensions in English: resolving the contradictory findings. *Language, 62*(2), 384.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1992). On the complexity of discourse complexity: a multidimensional analysis. *Discourse Processes*, *15*, 133-163.

Biber, D., & Clark, V. (2002). Historical shifts in modification patterns with complex noun phrase structures: How long can you go without a verb? In T. Fanego, M. J. Lopez-Couso, & J. Perez-Guerra (Eds.), *English historical syntax and morphology* (pp. 43-66). Amsterdam: John Benjamins.

Biber, D., & Conrad, S. (2009). *Register, genre, and style* (Cambridge textbooks in linguistics). New York: Cambridge University Press.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multi-dimensional comparison. *TESOL Quarterly*, *36*, 9–48.

Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics*, *15*(2), 223-250.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2-20.

Biber, D., & Gray, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL iBT® test: A lexico-grammatical analysis. *ETS Research Report Series, 2013*(1), I-128.

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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. New York: Longman.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. Oreilly Media.

Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds), *Dimensions of L2 performance and proficiency – investigating complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42–65.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

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., Kyle, K., Allen, L., Guo, L., & McNamara, D. (2014). Linguistic microfeatures to predict L2 writing proficiency: a case study in automated writing evaluation. *Journal of Writing Assessment*, *7*, 1-16.

DeClercq, B. D., & Housen, A. (2016). The development of morphological complexity: a cross-linguistic study of L2 French and English. *Second Language Research*, *35*(1), 71-97.

Deng, Y., Lei, L., & Liu, D. (2020). Calling for more consistency, refinement, and critical consideration in the use of syntactic complexity measures for writing. *Applied Linguistics*, 1–9.

Fang, Z., Schleppegrell, M., & Cox, B. (2006). Understanding the language demands of schooling: nouns in academic registers. *Journal of Literacy Research*, *28*(3), 247-273.

Ferris, D. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, *28*(2), 414-420.

Flahive, E., & Snow, G. (1980). Measures of syntactic complexity in evaluating ESL compositions. In J.W.Oller & K. Perkins (Eds). *Research in language testing* (pp.171-176). Rowley, MA: Newbury House.

Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition*, *18*, 299-324.

George, M., & Chomsky, N. (1963). Finitary Models of Language Users. In *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). New York: Weily.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, *9*(2), 123-145.

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.

Hall, G. (2011). *Exploring English language teaching: language in action*. London: Routledge.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar*. London: Routledge Taylor & Francis Group.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston: Heinle & Heinle.

Hawkins, J. (2004). *Efficiency and complexity in grammars*. New York: Oxford University Press.

Housen, A., & Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, *30*(4), 461–473.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Philadelphia: John Benjamins Pub.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels. NCTE research report no. 3.* Champaign, IL: National Council of Teachers of English.

Kyle, K. (2016) *Measuring syntactic development in L2 writing: fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Doctoral Dissertation).

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., 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*, 1-13

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, *12*, 439–448

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.

Macdonald, P. L., & Gardner, R. C. (2000). Type I Error Rate Comparisons of Post Hoc Procedures for I j Chi-Square Tables. *Educational and Psychological Measurement*, *60*(5), 735–754.

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, *29*, 3–15.

Musgrave, J., & Parkinson, J. (2014). Getting to grips with noun groups. *ELT Journal*, *68*(2), 145–154.

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.

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.

Ortega, L. (2003). Syntactic complexity measures and their relationship with L2 proficiency: a research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492-518.

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, *29*, 82-94.

Park, H. (2017). *Syntactic complexity as a predictor of second language writing proficiency and writing quality* (Doctoral Dissertation).

Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly, 14*(1), 61.

Pustejovsky, J., & Stubbs, A. (2013). *Natural language annotation for machine learning* (2nd ed.). Sebastopol, CA: O'Reilly Media.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*(1), 160–164.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language, 30,* 3–26.

Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing, 23*, 497–519.

Sebeok, T. A. (1960). *Style in language*. Cambridge: Technology Press of Massachusetts Institute of Technology.

Sharpe, D. (2015). Your chi-square test is statistically significant: Now what? *Practical Assessment, Research & Evaluation, 20*(8), 1-10.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: a longitudinal case study in the acquisition of finnish. *Applied Linguistics, 31*(4), 532–553.

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33*(2), 149-183.

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing, 32*, 17-35.

Tucker-Mclaughlin, M., & Hubbard, G. (2013). Making It Clear. *Electronic News*, *7*(2), 55–71.

Vyatkina, N. (2013). Specific syntactic complexity: developmental profiling of individuals based on an annotated learner corpus. *Modern Language Journal*, *97*(S1), 11–30.

Wang, S., & Beckett, G. (2017). "My excellent college entrance examination achievement" noun phrase use of Chinese EFL students' writing. *Journal of Language Teaching and Research*, *8*(2), 271-277.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

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