EXAMINING SYNTACTIC COMPLEXITY IN EFL ACADEMIC WRITING

LIMING LIU

Ph.D

The Hong Kong Polytechnic University

2018

The Hong Kong Polytechnic University

Department of English

Examining Syntactic Complexity in EFL Academic Writing

Liming Liu

A Thesis Submitted in Partial Fulfilment of the Requirements for

the Degree of Doctor of Philosophy

April 2017

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Liming Liu

ABSTRACT

The sophisticated nature of academic knowledge building and argumentation necessitates equally complex and elaborated linguistic representation. This has become one of the biggest linguistic challenges for second language (L2) writers of English for academic purposes (EAP). This study investigates syntactic complexity of advanced L2 academic writing, with SC understood as the sophistication and variety of meaning-making linguistic devices available in the grammar of EAP. The study takes a contrastive corpus-based approach, comparing various dimensions of SC between EFL dissertations written by Chinese postgraduate students and published research articles. The aim of the comparisons is to identify areas of syntactic complexity where EFL student writers lag greatly behind expert writers, which can be pedagogically taken up in EAP instruction. The study first employs commonly used L2 syntactic complexity measures to see which measures show significant differences between student writers and expert writes. Subsequently, two major dimensions of syntactic complexity unique to academic writing are examined in detail: noun phrase complexity and the complexity of clause combination. Although register variation research has revealed that academic discourse is characterised by noun phrase complexity while clause combination is typical of everyday conversation, it is argued that these two syntactic complexity dimensions represent different functional and semantic complexification in academic writing: while the former condenses meanings of processes, relations, and attributes, which are canonically construed as clauses**,** into noun

phrases, the latter encodes logical and semantic relationships between statements for the purpose of building arguments.

A number of findings stand out from the results. It has been found that student writing shows significantly lower scores on most syntactic complexity measures than expert writing. The analysis of noun phrase complexity shows that although students are similar to expert writers in selecting structural types for postmodifiers, the complex noun phrases they use are much shorter and contain fewer layers of multiple postmodification compared with expert use. Student writers also do not tend to use complex noun phrases in subject position in ways similar to expert writers. The analysis of finite adverbial clause combination reveals that student writers use fewer concessive and causal clauses than expert writers overall but that it is not uncommon to find instances of student writing where clause combination is unconstrained and obstructs effective communication. Student writing also contains fewer participle adverbial clauses, but cases of sentence-internal overuse of this clause type are spotted.

Attempts have been made to interpret the findings from functionally related perspectives based on careful textual analysis. Students' weak use of multiple postmodification in the noun phrase, besides indicating a lack of awareness for the function of this grammatical feature in EAP, also points to their underdeveloped disciplinary knowledge and participation needed to explicate entities and processes related to their research. On the other hand, inadequate concessive and causal clause combination indicates students' unpreparedness and lack of pressure to present justification for and potential external criticism of their statements and claims. Underuse of participle clauses not only indicates students' unfamiliarity with this clause-combining device but also their

unawareness of its discourse-coordinating and style-enriching function. In addition, possible transfer from Chinese rhetorical traditions is invoked to interpret students' weakness in both types of complexity. Theoretical contributions to L2 syntactic complexity research and implications of the findings of the study for EAP pedagogy in EFL contexts are discussed as well.

PUBLICATION ARISING FROM THE THESIS

Parts of this thesis have been published in an international refereed journal, as listed below. These parts include the data used for this study (Chapter 3), as well as the methodology (Chapter 3), results and discussions for syntactic complexity measurement (Chapter 4) and the study of complex noun phrases (Chapter 5). Due notes have been made throughout this thesis where content is based on the said journal article.

Liu, L., & Li, L. (2016). Noun phrase complexity in EFL academic writing: A corpus-based study of postgraduate academic writing. *The Journal of Asia TEFL, 13*(1), 48-65. doi:10.18823/asiatefl.2016.13.1.4.48.

ACKNOWLEDGEMENTS

I am grateful to many people and institutions for their contribution to my doctoral research.

I am deeply indebted to Dr Li Lan, my first chief supervisor, for guiding me through most stages of my studies. Coming from a Chinese background and educated in the UK for many years, Dr Li knows perfectly what a home-grown Chinese student like me needs to go through to become apprenticed to international academia. Working with her enormously transformed me academically, professionally, and ideologically. I have never failed to be impressed and inspired by her energetic, upbeat, considerate, and diligent attitude and approach to teaching, to research, and to life. I am thankful for the research assistantships kindly offered by Dr Li after I had run out of stipend after my normal period of study. Although Dr Li could not continue her supervision for my thesis after her retirement, I was blessed she made great effort to help create a board of examiners for my thesis.

My most special and respectful gratitude goes to late Prof Stephen Evans. After Dr Li retired, I had to have my chief supervisor changed, which had been worrying me for quite a while. Stephen's extremely generous offer to be my chief supervisor helped me immensely. He read and made detailed comments on my work with care and efficiency. In fact, only a few days before his fatal relapse, he sat down with me to discuss my thesis and show me how to revise it in every possible detail. I have no idea what I should do to compensate for the enormous and selfless effort he put in my thesis – I promise I will work this out.

I also wish to extend my sincerest thanks to Prof Kathleen Ahrens, my current supervisor. After Stephen's devastating passing, Kathleen generously came in to help me out for my final stages of study. She went through my work with amazing quickness and made detailed and incisive comments for my thesis to meet examination standards. I am grateful for the many intense meetings Kathleen arranged during this brief supervisory period, which greatly helped me tackle a number of small but thorny issues before my final submission. Also much to my surprise, she has advised me a lot on my future academic and professional career, and offered valuable comments on my job application materials. It is through her encouragement and advice that I have built more confidence in my value as an up-and-coming young scholar and teacher.

I would also like to thank Dr Foong Ha Yap, my co-supervisor, for the many meetups I had with her over lunch and dinner, in AG434, in the hallway, and over the telephone. I have benefitted greatly from Foong Ha's time and input and enjoyed our chats immensely. A devoted linguist on many Asian and European languages, Foong Ha has greatly enriched my understanding of how languages work over and above English which I mainly work on. She also advised and helped me a lot when I made difficult choices regarding my career path.

I am also thankful to Dr Yongyan Li of The University of Hong Kong and Dr Vincent Ooi of National University of Singapore for agreeing to act as external examiners of my thesis, out of their busy work schedules.

I also wish to thank all the PolyU teachers who have taught and helped me, especially those in the Department of English. I would like to thank Prof Martin Warren for teaching me discourse analysis and corpus linguistics and for agreeing to be the BOE

chair for my thesis. I would also like to thank Prof David Qian for teaching me research methodology and telling me a lot of fun anecdotes in academia, Prof Hans Ladegaard for kindly helping me out on two crucial supervisor changes, Dr Yao Yao for examining my PhD confirmation report and teaching me computational linguistics, and Dr Victor Ho for his earnest and encouraging suggestions on my finishing the thesis.

The other people I would like to thank in the Department of English are all the support staff, particularly Ms Vinci Leung, who have been providing all the research students with top-notch administrative support over the years with an efficiency and care that I was not used to before I came to PolyU. I would also like to take this opportunity to thank Hong Kong's University Grants Committee for offering me a monthly stipend during my three-year normal study period, without which it would have been impossible for me to embark on my doctoral research. I also feel blessed and privileged to have free access to the superb resources and facilities generously provided by the Department of English, making it the envy of many other departments and universities across the world.

I am thankful to my fellow research students and research staff, for their long-time help, company, encouragement, comfort, and all the goodies and delicacies they shared with me. AG415, where most of the research students work, has become a perfect retreat for both study and refreshment.

My thanks shall also go to my dear sisters and brothers at Tsim Sha Tsui Mandarin Baptist Church for helping and guiding me spiritually, culturally, and gastronomically over my years in Hong Kong. I have been blessed to have their company and teaching whenever I seem to be on the downside of life. I am also uniquely obliged to my other spiritual

soulmate Sinae Sim for her unofficial supervision of my progress with great patience, care, and agape.

Finally, it goes without saying I shall express my most heartfelt gratitude to my parents in Tianjin—my mom Jianying Li and my dad Zhi'an Liu—who have always been extremely supportive in every aspect of my life and study and whose unfailing love and care, though sometimes unvoiced, have been an unequivocally tremendous source of impetus for my determination to finish my PhD.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1. Purpose of the Study

This study examines the syntactic complexity of advanced EFL (English as a foreign language) academic writing. Taking a contrastive corpus-based approach, it compares a range of syntactic complexity features between Master of Arts (MA) dissertations and journal articles in the field of applied linguistics.

As a point of departure, syntactic complexity measures developed in L2 writing research are compared between student writing and expert writing to see what measures display significant differences between the two groups of writers. Subsequently, the study examines in detail two important dimensions of the syntactic complexity of academic writing: noun phrase complexity and the complexity of clause combination (or clausal complexity). A range of grammatical features typical of these two dimensions are compared and careful textual analyses of student writing are performed in order to qualitatively interpret the quantitative results. Instead of aiming to prove that student academic writing is weaker than professional academic writing in terms of syntactic complexity, the ultimate purpose of the study is to identify syntactic complexity features where students lag behind expert writers, which can then be taken up as areas for fruitful EAP instruction.

Additionally, since the study of syntactic complexity has long been separated from EAP writing research due to its lack of attention to functional and usage-based properties, this study also intends to explore possibilities where the functionally motivated features of syntactic complexity can be applied to EAP writing instruction.

1.2. Need for the Study

This study investigates the syntactic complexity of advanced EFL academic writing and addresses the issue by examining the dissertations written by Chinese postgraduate students. However, this topic does not easily fit into the overall research area of L2 syntactic complexity (e.g. Ortega, 2003; Wolfe-Quintero et al., 1998), which rarely involves such disciplinary genres as dissertation, literature review, and lab report written by L2 university students. Research on the syntactic complexity features of tertiary-level L2 writing has been predominantly based on student texts of basically homogeneous modes of discourse (Smith, 2003) in the form of short essays rather than associated with specific academic disciplines. These basic text types include narration, description, exposition, and argumentation, with argumentative essays being the most widely examined. Originating from the tradition of composition teaching in North America, these essay forms have become extensively used in ESL (English as a second language) and EFL programmes and in standardised and institutional English proficiency tests worldwide, such as the Test of English as a Foreign Language, or TOEFL (Crossley, Kyle, & McNamara, 2016; Raimes, 1990).

With topic prompts mostly derived from everyday social issues such as artificial intelligence and air pollution, essay writing has been generally conceptualised as academic because the purpose of a writing syllabus grounded in composing essays is to develop students' "core" academic literacy as a basis for writing requirements in academic disciplines. However, due to their generic limitation and their production in contexts unrelated to disciplinary academic activities, essays can be only marginally regarded as academic writing in its strict sense. By contrast, specific academic genres such as dissertations are produced for research (and examination) purposes and generically heterogeneous (i.e. multiple genres are at work), thus being able to be

conceptualised as academic writing proper. This is why the present study investigates MA dissertations as advanced academic writing, rather than the basic undergraduate essay types.

Therefore, the characterisation and theorisation of L2 academic texts have been almost exclusively based on non-disciplinary generic texts written for general intellectual purposes, with the examination of the syntactic complexity of academic texts written by L2 research students and scholars largely ignored. As the effects of genre and discipline on textual features increasingly come to be realised and emphasised, it is necessary for L2 texts written for specific academic purposes to take centre stage of research on L2 academic writing.

Although there has been a growing interest in research on EFL dissertation writing at the master's and doctoral levels, topics in this area are largely skewed towards such issues as organisational structure (e.g. Bunton, 1999; Kwan, 2006), difficulties and challenges facing L2 student writers (e.g. Bitchener & Basturkmen, 2006), authorial identity (e.g. Cadman, 1997), and course development (Flowerdew, 2015). To date, there has been little research on the syntactic complexity of EFL dissertations, despite long-standing interest in examining grammatical features of undergraduate essays.

The dearth of research on the syntactic complexity of EFL dissertation writing could be attributed to lack of dialogue between different areas of second language writing scholarship. The study of syntactic complexity of second language writing has been largely taken up by second language acquisition (SLA) researchers interested in developing syntactic complexity measures to track L2 writing development across distinct curriculum or proficiency levels (Ortega, 2015). Not solidly based on functional and semantic dimensions of syntactic complexity, measures used in SLA studies are

usually derived from researchers' intuition of what constitutes complex syntax (Biber, Gray, & Poonpon, 2011). In this paradigm, syntactic complexity has been typically thought of as long production units and elaborated clause combination, which has generated numerous—although often overlapping—measures tapping into these two dimensions (Ortega, 2003). One of the problems of these measures is that they are at best broad-brush structural representations of what is involved in long production units and clausal complexity, leaving the functional and cognitive basis of syntactic complexity largely unaccounted for. In fact, the development of such measures mainly suits the purpose of writing assessment and building acquisition models for general L2 writing growth rather than to provide practical pedagogical implications for teaching academic writing in specific educational and disciplinary contexts.

On the other hand, research on dissertation writing is mainly taken up by EAP scholars who are directly concerned with designing and/or teaching academic writing courses to research students from specific academic disciplines (Paltridge, 1997, 2002; Richards, 1988). Examination of linguistic features in EAP contexts usually takes account of their discoursal and contextual/disciplinary correlates, seldom carried out in isolation as is the case with generalised L2 writing research (Cadman, 1997; Starfield & Ravelli, 2006). Syntactic complexity measures as purely structurally conceptualised would not be seen as practically conducive to the teaching of academic writing. From another perspective, EAP researchers may not deem it challenging for research students to produce syntactically accurate and complex texts considering their advanced academic attainment, hence their indifference to the study of syntactic complexity in examining dissertation writing.

However, the gulf between SLA-based syntactic complexity research and EAP-based academic writing research has recently seen some signs of narrowing as EAP

researchers have begun to embrace functionally motivated perspectives of syntactic complexity (e.g. Parkinson, 2015; Parkinson & Musgrave, 2014). On the other hand, SLA researchers have also paid increasing attention to the functional and usage-based underpinnings of syntactic complexity (e.g. Ortega, 2015). However, more work needs to be done to situate the study of syntactic complexity in more generically and disciplinarily significant academic contexts. Being at an intersection of SLA and EAP, this study sets out to intensify this burgeoning interaction by studying the syntactic complexity of EFL academic writing.

1.3. Significance of the Study

This study investigates various dimensions of the syntactic complexity of advanced EFL academic writing. Findings of the study have both theoretical contributions and pedagogical implications. Theoretically, the study contributes to syntactic complexity research in second language writing by going beyond largely structurally-motivated syntactic complexity measures, shifting the focus to their functional and usage-based underpinnings and examining student writing in close association with these meaning-making rationales.

Another theoretical concern about syntactic complexity relates to the relationship between noun phrase complexity and clausal complexity, which has been predicted to develop in an opposite manner along with proficiency and literacy growth (Ortega, 2003). That is, clausal structures are predicted to decrease as proficiency increases, gradually replaced by greater use of information-packing complex noun phrase structures. With the two writer groups in this study representing two distinct literacy levels, comparing these two dimensions of syntactic complexity may also help examine their predicted relationship and relative importance in academic writing.

On the pedagogical side, although EAP practitioners and researchers have taken an increased interest in postgraduate dissertation writing (Lee & Casal, 2014; Leki, Cumming, & Silva, 2010; Xie, 2016), little attention has been paid to students' problems in coming to terms with syntactic complexity to cater for the complex meaning-making needs of academic writing. Since lacking syntactic variety and sophistication may undermine the writer's academic credibility as a capable participant and discussant in the research domain, findings of the study that reveal specific syntactic weaknesses in student writing should be able to raise students' awareness of its endemic textual features and help teachers better target those weaknesses in pedagogical practices.

1.4. Research Questions

In order to achieve the research objectives stated above, this study comprises three independent and related studies addressing three overarching research questions. The first study compares the syntactic complexity measures between EFL academic writing and expert writing; the second study compares grammatical features of noun phrase complexity; and the third study compares features of clause combination. Specifically, the three research questions are:

Research Question 1: Which syntactic complexity measures show significant differences between advanced EFL academic writing and expert academic writing?

Research Question 2: What are the grammatical features of noun phrase complexity where EFL student writers lag behind expert writers?

Research Question 3: What are the grammatical features of clause combination where EFL writers lag behind expert writers?

1.5. Methodological Preview

The three research questions are respectively addressed in three independent yet interrelated investigation endeavours, presented in three chapters (see the following section for details). This study takes a corpus-based contrastive approach to examining the syntactic complexity of advanced EFL academic writing, with expert academic writing serving as the frame of comparison. For EFL academic writing, a corpus of MA dissertations in applied linguistics is built; a comparative corpus of published research articles in the same field is compiled. In this study, MA dissertations are defined as advanced EFL academic writing. This differs from the general practice in L2 writing research which usually employs undergraduate essays to represent advanced EFL writing. The decision is made on the argument that basic undergraduate essays types such as description and argumentation can hardly be taken as "advanced" academic writing. Arguably, they should be thought of as "less advanced" than dissertations and full-length articles for research purposes. The discipline applied linguistics is selected mainly because of data accessibility. On the Chinese mainland, only students from English major programmes write up their MA dissertations in English and the discipline applied linguistics is mostly offered to these students.[1] On another account, applied linguistics is also an area that the present author is familiar with, which is conducive to making informed decisions in data collection.

The three investigations are enlightened by different areas of L2 writing scholarship. The first investigation draws on research into L2 syntactic complexity measurement in second language acquisition, relying on quantitative measures to

[1] English major programmes offered by Chinese universities aim to produce professionals in a range of English-related occupations. Programmes at the undergraduate level are mainly focussed on language proficiency courses with a small research component and master-level programmes are developed on the basis of academic disciplines, such as translation, literature, theoretical linguistics, and applied linguistics.

answer Research Question 1. The other two investigations are based on theoretical discussions and empirical findings regarding the distinctive linguistic features of academic discourse, subjecting them to frequency comparisons; works of functional linguists and corpus linguists on register variation form the analytical basis.

1.6. English as an Academic Lingua Franca

As English has become the world's academic lingua franca (or language for wider communication), it has greatly impacted on educational practices and policies related to the medium of communication in academia (Coleman, 2006; Evans & Morrison, 2011, 2016; Hu, 2009). Two areas in higher education that have been affected by this globalised role of English are the medium of instruction and academic publication. While English as the medium of instruction (EMI) has long been practiced in ex-colonial societies where English is one of the official languages (i.e. ESL social settings), social contexts in which it is only a foreign language (i.e. EFL settings) have also begun to embrace this policy. In fact, EMI has been gaining considerable momentum in continental Europe (Coleman, 2006) and its implementation in the Asia Pacific is also on the increase, for example in South Korea (Kang, 2012) and on the Chinese mainland (Hu, 2009).

An important driver for implementing EMI in higher institutions across the globe is the need and pressure to publish in international journals, as most of them use English as the language of publication (Hyland, 2009, p. 5). Because participation in international publication serves as a major means to enhance the prestige and credibility of institutions and individuals, the extent to which English can be used for this purpose determines academic success to a great degree. As the lingua franca has become the

language of knowledge making and dissemination, a lack of it means almost complete inaccessibility to international academic communication.

For both EMI and international publication, writing in English for academic purposes (EAP) constitutes a substantial component. Using English as the medium of instruction indicates that subject essays, lab reports, and dissertations as crucial forms of assessing students' academic achievement have to be completed in English.[2] While academic writing is a challenging activity for both native speakers (NSs) and non-native speakers (NNSs) alike, presenting complex description and argumentation in a second language seems a far more daunting task. Even in societies where EMI has a long history, for instance in Hong Kong, EAP writing has proved to be problematic for local students, making EAP courses a crucial facilitation for educational success (Evans & Green, 2007). By contrast, NNS students studying in EFL contexts where only limited EMI is provided no doubt need more intensive pedagogical facilitation if they are required to perform effective EAP writing.

In EFL societies such as China, the significance of EAP writing seems greater for higher degree research students than for undergraduates. A good number of the former group look to find research or teaching positions in universities or research institutes after graduation, so publication experience in international refereed journals has increasingly become mandatory for job opportunities in higher education nowadays. As noted by Hyland (2009, p. 180), "[m]any prestigious Chinese universities stipulate that their PhD students must have at least one paper accepted by an international journal before they can graduate". However, this stipulation mainly applies to doctoral students in academic disciplines in the natural and engineering sciences (Li, 2007, 2016) as

[2] This study uses the term dissertation rather than thesis to refer to the written work required of master's degree candidates.

scientific writing per se in these disciplines is relatively easier due to its heavy reliance on rigorous and convincing mathematical extrapolation and experimental proof. In Li's (2006) in-depth case study of a Chinese physics doctoral student's strenuous effort to get published in a renowned English-language journal, the student scholar's challenges to successful publication mainly come from methodological and theoretical issues rather than from imperfect academic writing quality—although his writing has been substantially polished by multiple individuals before final publication. Publication for these students is further eased by the fact that research in the hard sciences is typically conducted through collaboration between professors, research staff, and students (Baldwin & Austin, 1995); therefore, publications are usually multi-authored and students are not on their own.

In contrast, writing in the humanities and social sciences requires reasoned argument and explicit interpretation (Hyland, 2006), making international publication much more challenging. Again, since the nature of research in the soft sciences is often solo, the student is not part of a research team and has to work on her own for research writing and publication. Therefore, strong pedagogical support needs to be provided to students in these areas who look to develop their EAP writing skills for future participation in international publication. A good understanding of the linguistic problems in students' texts constitutes a solid prerequisite for effective pedagogical support for EFL students' academic writing. This study is devoted to such a descriptive and explanatory programme.

1.7. Preview of Linguistic Features of Academic Discourse and Syntactic Complexity

A large body of research has been done to examine the linguistic features of academic discourse, with efforts made to identify the surface linguistic features and their correlations with discourse functions. This line of research has been mainly undertaken by function-oriented textual linguists, whose work has provided valuable information and insights for teaching academic writing. Based on this work, Hyland (2006, pp. 13-14) summarises the linguistic features of academic discourse as high lexical density, high nominal style, and impersonal constructions. Given that the linguistic landscape of a text is largely discourse-driven, writers' preferences for lexicogrammatical devices also need to be understood from discourse-related perspectives, such as academic discipline, communicative purpose, and readership. For example, since an important aim of research writers is to create an image of objectivity and impartiality in their writing, it is things and entities rather than people that participate in the unfolding of discourse, which can best be construed through linguistic devices connected with the nominal and impersonal style. More detailed illustrations of form-function mappings in academic discourse will be given in Chapter 2.

One of the consequences inherent in the abovementioned features, particularly high lexical density and nominal style, is what can be called linguistic complexity. The notion of complexity can be employed to characterise complex textual quality on the part of the writer and its effect on text comprehension on the part of the reader. The need for academic discourse to be textually and cognitively complex is largely determined by the fact that complex concepts, processes, and argumentations have to be coded into explicit expression. Text complexity is a reflection of the complexity of the subject matter being written about and the research being undertaken along with the

author's knowledge and interpretation of the whole research process and results. Here the author's level of expertise plays a crucial role in the creation of linguistic complexity considering that topic familiarity has a significant effect on text quality, including critical thinking and syntactic complexity (Stapleton, 2001; Stapleton & Wu, 2015; Tedick, 1988). To some extent, elaborated textuality adds to the credibility of the research writer as an eligible participant in a particular research domain.

However, syntactic complexity should be commensurate with the need to express complex information and ideas, and should not serve as an *a priori* objective of L2 writing. Students need to be warned against the deliberate pursuit of using complex lexis and grammar to create an "advanced" style which is an unfit and awkward mismatch with underdeveloped content knowledge and control of grammatical resources. In some extreme cases, college ESL teachers would even make it mandatory for students to write short sentences to avoid disfluency and infelicities. While it is justifiable in academic writing to use elaborated language for complex description, evaluation, and argumentation, information-integrating devices such as premodification and postmodification are prone to disorganised abuse. Therefore, it would be always necessary for EAP writing courses to be wary of the downsides of ineffectively deploying syntactic complexity features.

In second language acquisition research, syntactic complexity usually refers to the sophistication and variety of grammatical structures used in writing tasks (Wolfe-Quintero, Inagaki, & Kim, 1998).[3] Researchers have developed a range of measures for examining L2 writers' syntactic complexity. Norris and Ortega (2009) identified five types of syntactic complexity measures after reviewing empirical studies from the

[3] Syntactic complexity is also called grammatical complexity by some scholars, especially Douglas Biber and colleagues, as will be mentioned in the literature review chapter; the two terms are interchangeably used in this study.

1960s to 2000s. They found two types of measures frequently employed for measuring the complexity of L2 writing: 1) measures based on the length of production units, e.g. mean length of clause and mean length of T-unit;[4] 2) measures based on subordination, e.g. clauses per sentence. The dominance of these two types of metrics can be accounted for by researchers' intuition that longer production means better fluency and that more subordinate structures result in elaborated reasoning and argumentation.

However, length measures and subordination measures sometimes return inconsistent patterns on syntactic complexity. Studies employing measures that gauge different dimensions of complexity may produce conflicting results. It can be difficult for one to interpret these results without a good knowledge of different types of syntactic complexity. To better inform syntactic complexity measurement, Ortega (2003) and Norris and Ortega (2009) have made a case for a multi-dimensional description of syntactic complexity, involving the measurement of subordination, length, and phrasal structures. The incorporation of noun phrase structures into the general conceptualisation of syntactic complexity comes as a result of the research domain's increased awareness of the different underpinnings and functions of distinctive linguistic features of academic discourse, thanks to the research of functional and corpus linguists interested in register and genre variation. Therefore, recent developments in L2 syntactic complexity research have shifted its focus to tapping into the functional, cognitive, and usage-based representations of syntactic complexity (Ortega, 2015).

[4] T-unit, or "terminable unit", is a term proposed by Hunt (1965) referring to an independent clause (i.e. main clause) together with any number of subordinate clauses appended to it. For example, both *I went to bed* and *I went to bed after I brushed my teeth* are T-units.

1.8. Organisation of the Thesis

Being an empirical examination of syntactic complexity of EFL academic writing, this thesis includes the following major components, divided into seven chapters. This Chapter, i.e. Chapter 1, has introduced the emerging importance of EAP writing for tertiary-level students studying in EFL contexts and the lack of research into academic text types beyond the basic examination-oriented student essay, thus making a case for examining the linguistic features of EFL research students' dissertations as a more meaningful genre of academic writing. Chapter 2 presents a comprehensive review of literature on a range of relevant themes, including the conceptualisations of writing as product and as process, the notion of genre, different understandings of grammatical units such as clause and subordination across schools of linguistic analysis, and research on dissertation writing, syntactic complexity of academic discourse in general, and syntactic complexity of second language writing. The literature review establishes two major dimensions of syntactic complexity to be carefully studied in this thesis: noun phrase complexity and clausal complexity. Chapter 3 explicates the methodology and analytical procedures for examining different dimensions of syntactic complexity: this study adopts a contrastive corpus-based approach to the study of EFL academic writing, using published expert writing as a frame of standard reference. Three independent, but related, studies are designed to examine overall syntactic complexity, noun phrase complexity, and clausal complexity, respectively. Chapters 4, 5, and 6 present the results of the three empirical studies and discuss the results in relation to their syntactic, semantic, discourse, and cultural correlates. Chapter 7, the final chapter, concludes the thesis by summarising the major findings, discussing theoretical contributions and pedagogical implications, considering analytical limitations, and making recommendations for future research.

CHAPTER TWO

LITERATURE REVIEW

This chapter provides a comprehensive review of literature on themes relevant to the examination of syntactic complexity of EFL academic writing, starting from general areas of enquiry to specific ones. It aims to locate the major disciplines and research domains having direct implications for studying EFL academic writing and to identify recent developments and existing inadequacies in research on L2 syntactic complexity.

In Section 2.1, the review sets out to present the overall theoretical and disciplinary framework within which this study operates, including contrasting views of the nature of writing, fields of enquiry related to L2 writing, and different schools of grammatical analysis. A review of these topics helps to lay the overall ideational and disciplinary basis for this study. In Section 2.2, the notion of genre as providing essential theoretical orientations to textual analysis is brought up to inform the present study. In Section 2.3, the chapter goes on to examine research on dissertation writing and tries to discover its trending and inadequate lines of enquiry. In what follows, the central topic of this study, i.e. syntactic complexity, is carefully reviewed. Section 2.4 focuses on theoretical and empirical studies of syntactic complexity of academic writing by scholars in discourse analysis and corpus linguistics and Section 2.5 examines empirical research on syntactic complexity development in the field of second language acquisition, with Section 2.6 reviewing studies on the syntactic features of Chinese EFL students. In so doing, it is expected that both common grounds and gaps between linguistic descriptions and empirical investigations concerning EFL syntactic complexity can be found to make a case for the present project.

2.1. Preliminaries

2.1.1. Writing as Product versus Writing as Process

Being a textual analysis of EFL academic writing, this study adopts a methodologically product-oriented, rather than process-oriented, approach to writing instruction and research (Grabe & Kaplan, 1996; Hyland, 2009), while fully recognising that surface features of a text are strongly driven by such discourse-related factors as communicative purpose, author identify, and topic familiarity. A product approach to writing research focuses on examining *texts* as the product of writing, either their clause and sentence level linguistic forms or discourse level structures and processes. However, this does not mean that texts are approached as static, autonomous objects to be described and analysed independently of contexts, writers, and readers. Instead, an ideal product-approach is devoted to explicating the linguistic and discourse features of specific text types and attempts to explain these features in terms of cognitive-semantic and sociocultural factors that contribute to their development. Thus, it rejects the idea that just replicating surface linguistic and discourse features without considering their functional and social underpinnings is enough for writing development.

Analyses of forms and structures for pedagogical purposes need to be constantly related to ways they help construct meaningful and effective texts. As will be seen later in this chapter, research on syntactic complexity of student writing (in both L1 and L2) has largely taken a form-focused approach that somewhat ignores the "functional and usage-based rationales" of certain syntactic features (Ortega, 2015, p. 83). However, in this study these rationales will be frequently picked up in justifying the selection of linguistic features to be examined and explaining their representation in student writing. Such an approach fits in with the pedagogical principle *focus on form* (FonF) in second

language acquisition (Long, 1991), which stresses the importance of form-function correspondence for awareness raising to take place.

However, it needs to be noted that the product approach has been under strong criticism from process-focused approaches to writing teaching and research. Highly influential for college English writing courses in North America and some other parts of the world, process approaches dismiss teaching grammatical features as not capturing important process characteristics such as writers' personal creativity, cognitive processes, and immediate context (Hyland, 2009, p. 18). Under this approach, teachers are only encouraged to stimulate students' thinking about the planning and idea development of writing by virtue of a range of tasks. Correspondingly, students are expected to break away from fixed writing models and linguistic formulae evident in others' texts.

Research on writing processes usually takes qualitative, *emic* approaches such as think-aloud protocols, retrospective interviews, and non-intervening observation to reveal what writers think and do during writing. However, such methods, particularly think-aloud protocols, have been criticised for offering only incomplete and even distorted pictures of complex cognitive processes, making results thus obtained not reliable. In addition, process-based research into L2 writing has yielded disappointing results about writing processes, merely supporting common-sense intuitions about qualities of good writers (see Silva, 1993). Indifference to how writing processes are linked to linguistic features reduces the possibility of effectively tracking students' written output development over time. Therefore, this study does not base itself on the epistemological view that effective writing is only determined by factors involved in the writing process. However, writers' cognitive processes that have explanatory

strengths for certain linguistic features of student writing will be taken into account in due course.

2.1.2. Disciplinary Framework

Guided by the writing-as-product approach, this study draws on concepts and methodologies from a number of sub-disciplines in the broader field of applied linguistics, including English for academic purposes (EAP), second language writing, discourse analysis, genre analysis, and corpus linguistics. Some of these are methodological tools for analysing linguistic data but at the same time can be regarded as theories of language， including genre analysis, discourse analysis, and corpus linguistics; others concern the teaching of English language skills in higher education contexts and the analysis and evaluation of students' academic writing.

First, research into postgraduate dissertations originates from an interest in examining this particular text type in its own right (Leki, Cumming, & Silva, 2010). The concept of *genre*, a major topic in EAP, applies here since dissertations differ from other types of academic discourse (e.g. research articles and scholarly monographs) in the social actions they perform and communicative purposes they are to achieve (Bunton, 1998). In this sense they need to be understood as having different generic features at the lexicogrammatical and discourse levels as compared with, say, the research article. However, as will be discussed below, dissertations and research articles also have much in common as the former is usually modelled on the latter within a particular discipline, particularly in terms of syntactic features.

Second, the authorship in focus in this study is limited to English as a foreign language (EFL) writers at a highly advanced literacy level. This naturally relates the present study to the vast body of research into second language writing, under the

broader rubric of second language acquisition (SLA). SLA research has taken a keen interest in examining the complexity, accuracy, and fluency of L2 writing and speech at a particular curriculum or age level or their development over time (Lu, 2010; Ortega, 2003). Among the three variables, syntactic complexity is the most widely researched. However, research on L2 dissertation writing has proceeded more from the perspectives of generic structure (e.g. genre analysis of the moves and steps of different sections of the dissertation), rhetoric (e.g. stance-taking and argument building), and other sociolinguistically oriented notions (e.g. authorial identity) than from the angle of syntactic complexity. This gap might have been caused by the belief that academic writing at the postgraduate research level is no longer so much concerned with expanding text complexity as with dealing with the textual and interpersonal aspects of the dissertation. Therefore, this study attempts to address this gap by putting EFL dissertation writing in the perspective of syntactic complexity as often taken up by L2 writing research.

Finally, corpus linguistics, converging on the contribution from theoretical linguistics, computational linguistics as well as artificial intelligence, has developed into a commonly employed methodology in many areas of applied language sciences. A major concern for empirical research is sample size and representation (Biber, 1993), the magnitude of which has significance for the degree of result trustfulness and generalizability—in other words, the larger and comprehensive the sample, the more representative of reality it is. It is precisely in collecting and automatically analysing huge amounts of naturally occurring language data that corpus linguistics excels. Due to its ability to computationally process and compare immense language data, corpus linguistics has made substantial contribution to the revelation of linguistic features varying across different registers (e.g. Biber, 1988), particularly between academic

writing and informal speech. As will be seen later, many of the observations by linguists and grammarians on lexicogrammatical features of academic writing have been empirically confirmed and complemented by corpus linguistics research.

As such, the review of literature will cover the three areas so far discussed as having implications for the issue of syntactic complexity of academic writing in general and EFL academic writing in particular.

2.1.3. Notes on Grammatical Frameworks for Textual Analysis

The grammatical framework used for textual analysis in this study includes traditional descriptive grammar of English (e.g. Biber, Johansson, Leech, Conrad, & Finegan, 1999; Quirk, Greenbaum, Leech, Svartvik, & Crystal, 1985) and functional linguistics (e.g. Halliday & Matthiessen, 2004). Although largely based on Quirk et al.'s framework, Biber et al. (1999) have also employed concepts and analytical methods from theoretical linguistics, including transformational generative grammar (e.g. *raising* in Ch. 9) and dependency grammar (e.g. *valency* in Ch. 5). This study will make constant reference to Biber et al. for frequency information on the use of certain linguistic features in academic discourse and other registers.

While not employed for data analysis, Halliday's systemic functional linguistics (SFL) will mostly be evoked for its insights into the multifaceted functions of linguistic forms. And this study will only borrow a small amount of its terminology since most of it is taken from the Prague school of linguistics and could be unfamiliar to scholars in educational linguistics. Nevertheless, important concepts of SFL on grammatical functions not available in or more revealing than traditional grammar will be taken up and explained where necessary. At this stage, a number of key concepts need some illustration to guide subsequent discussions.

2.1.3.1. Defining *Clause*

This study uses the term *clause* to refer to both finite and non-finite clauses, in line with the practice of the above grammatical frameworks. However, a dilemma occurs when it comes to measures of syntactic complexity in second language writing research (see also Yang, Lu, & Weigle et al., 2015, p. 54). It has been the tradition in L1 and L2 writing research to count only finite clauses in clause-related measures, with non-finite structures regarded as variants of verb phrases. As will be seen in more detail later, such measures include length measures (e.g. mean length of clauses) and ratio measures (e.g. complex nominals per clause), missing out on non-finite structures from such measurements and creating problems in functionally interpreting the results of measures using a grammatical framework with a broader scope of clauses. Functionally speaking, it would make more sense to take non-finite structures as clauses because they often enter into a semantic relationship with the main clause, especially with participle structures in an adjunct position, such as the following:

(1) ***Building on a study of six English as a second language (ESL) teachers' questioning behavior****, Long and Sato conclude that….*[5]

Although semantically somewhat indeterminate, the present participle adjunctive structure in the above example encodes something of a temporal or causal relationship with the ensuing main clause. Such a property better suits clauses rather than phrases, hence non-finite structures being regarded as clauses in this study. However, caution will be constantly exercised in relating syntactic complexity measures with functional claims about the role of clauses in academic discourse.

[5] Examples in this chapter are taken from the expert corpus used for this study (see Chapter 3).

2.1.3.2. Disentangling Clause Combination

Now that the scope of clause has been demarcated, it is necessary to deal with the situation in which more than one clause is included in a sentence. Two processes by which more than one clause is used to form grammatical sentences are recognised by traditional grammar: *coordination* and *subordination*. Of the two, coordination is relatively straightforward and subordination has created much complication. Coordination is the process by which two independent clauses are combined by a coordinator (e.g. *and*, *but*, *so*) to form a sentence and the two clauses share equal syntactic status, such as the following sentence with two coordinate clauses connected by *and*:

> (2) *On the other hand, students should understand that people's speeches and practices help to reproduce social inequalities and must always be brought into question,* ***and this understanding is not always reached through co-construction of meanings***[coordinate clause].

Although flexible in its realisations (Williams, 1978), the function of coordination shows a homogeneous picture, that is, to present two or more independent but related ideas. This way of clause combination is often called *parataxis* (adjective form paratactic) in functional linguistics (Givón, 2009; Halliday & Matthiessen, 2004) with a similar but rather broader connotation than coordination. In most cases, these two terms are interchangeably used in this study. By contrast, the broad term *subordination* involves a much wider variety of ways in which clauses combine or are attached to other grammatical categories, and different functions of "subordinate clauses" would emerge out of this variety. As will be seen, clause types that are traditionally defined as subordinate contribute to different processes of syntactic complexity. Therefore, it is necessary to distinguish these processes in the first place.

Two distinct ways of clause arrangement traditionally subsumed under the notion of subordination are distinguished by SFL: hypotaxis and embedding.[6] With *hypo-* meaning "beneath" and *taxis* "arrangement", hypotaxis is the grammatical arrangement of clauses of "unequal status…in which one dominates and the others are dependent on it" (Matthiessen, Teruya, & Lam, 2010, p. 114). In English, finite adverbial clauses (Ex. 3) and non-restrictive relative clauses (Ex. 4) are hypotactically combined with their main clauses, forming a range of "logico-semantic" relations (Halliday & Matthiessen, 2004, Ch. 7):

(3) *…the longitudinal nature of the study was explained,* ***although the exact linguistic focus of the study was not specified (to avoid test effects)***[finite adverbial clause].

(4) *Hypothesis 1 posited that the three-factor models (i.e. Models 2A-2D),* ***which were more constrained than the four-factor model (i.e. Baseline Model)***[non-restrictive relative clause]*, would not be significantly worse than the Baseline Model (Model 1)….*

These clause types are collectively called subordinate or dependent clauses in this study. By contrast, embedding is the process by which clauses are *rank-shifted* (downranked) to become clause elements: restrictive relative clauses become part of noun phrases (Ex. 5) while nominal clauses become subjects (Ex. 6) or complements (Ex. 7):[7]

(5) *…and the author was sitting somewhere behind them to note the actions* ***that could not be recorded by the software program***[restrictive relative clause]…

[6] In traditional grammar and formal linguistics, embedding and subordination are used to refer to the same concept.

[7] Also called complement clauses, nominal clauses can be used in syntactic positions suitable for nouns, i.e. as subjects, objects, and complements.

(6) ***How the students negotiate in the academic contexts to balance and counterbalance unequal power relationships***[complement clause as subject] *strongly influences their access to already limited host culture language resources.*

(7) *Consistent with* ***what is evident in Table 1***[complement clause as complement]*, there was a statistically significant overall priming effect….*

In such cases, there is no inter-clausal logico-semantic interaction. It is for these reasons that relative clauses and nominal complement clauses function to expand the clause rather than to enact clause combination, although more specifically the former joins the noun phrase and the latter independently serve as a nominal group (Matthiessen et al., 2010, p. 249). These clause types are collectively called embedded clauses in this study. Unfortunately, studies in second language writing, until quite recently, have not attempted to distinguish hypotaxis and embedding, with the consequence that an array of measures gauging different clause combining processes are used to represent a single concept—the complexity of subordination. This issue will be reviewed in detail later.

2.2. Degree Dissertations as a Genre

Derived from various traditions, the notion of genre has generated a complicated body of scholarship (Hyon, 1996). Hyon identified three traditions of genre relevant for ESL: 1) English for specific purposes (ESP), 2) systemic functional linguistics (SFL), and 3) North American new rhetoric studies. It has been found that ESP and SFL approaches to genre are concerned with revealing the linguistic features of written texts while the new rhetoric approach attempts to offer insights into instructional contexts for academic and professional genres. Therefore, it can be seen that the ESP and SFL approaches are of particular relevance for the present study in addressing the linguistic features of written texts for pedagogical purposes.

However, these two approaches differ in their text and educational focus, with ESP concerned with university-level academic discourse and SFL with elementary and secondary students' literacy development. Thus, genre theory in the ESP tradition would have direct implications for the study of advanced EFL academic writing, but SFL descriptions of genre are also essential to understanding how linguistic features of text types are shaped by and interact with their functional and sociocultural underpinnings. Accordingly, the review of genre will focus on these two theoretical traditions, beginning with ESP.

2.2.1. Genre and ESP

Since its inception in the seminal works of Swales (1981, 1990), genre has become an important topic in ESP research. A distinctive feature of the definition of this concept proposed by Swales (1990) relates to the communicative purpose(s) shared by users of a type of text within a discourse community. Communicative purpose entails a number of connected factors jointly constructing a genre, involving the identity and relationship of writers and readers, the immediate goal of production, access to readership, and text evaluation criteria. Academic writing has by far been the most researched area in ESP, within which the research article and postgraduate dissertation are two major genres under close examination.

The difference between master's and doctoral dissertations lies mainly in the scale of research and the issue of theoretical contribution (Madsen, 1992; Paltridge & Starfield, 2007). While a master's dissertation is usually 10,000 – 20,000 words in length, a doctoral dissertation in the humanities and social sciences needs to be 50,000 – 80,000 words long in line with the British and commonwealth convention of research

degrees. The doctoral dissertation also has to make an original theoretical contribution to existing knowledge (Madsen, 1992).

While research articles and dissertations have certain linguistic and rhetorical features in common, e.g. shared lexical and phraseological choices, syntactic attributes, the IMRD (introduction, method, results, and discussion) format, the two text types differ significantly in many respects (Paltridge, 2002; Thompson, 2012), e.g. readership, purpose, scale, and level of sophistication. Whereas the published research article is primarily aimed at disseminating and sharing up-to-date information with a wide scope of expert readership, the dissertation is a substantial, complex piece of work for assessment purposes demonstrating the degree candidate's knowledge of the discipline in question and their competence in reporting on large-scale independent research. These differences may even have impacts on micro-level linguistic features. For example, Swales (1990) pointed out that dissertations employ more metadiscourse than research articles due to the constant necessity to orient readers to different parts of the extended text. Elsewhere, a much greater degree of variation exists in the organisational format of the dissertation (Paltridge, 2002), but the research article mostly follows the traditional IMRD format. For example, dissertations in the humanities and social sciences are often structured in a topic-based format.

Despite the differences, dissertations and research articles share considerable similarities in terms of style and content. Both being academic discourse results in similar syntactic and lexical features. Multidimensional analysis of register variation (Biber, 1988) has found a set of grammatical features across five register dimensions that are uniquely shared by academic discourse, e.g. the use of nouns and nominalisations, prepositions, attributive adjectives, relative clauses, and conjuncts. An

important part of postgraduate research studies is to master the syntactic and lexical features characteristic of the genre of general academic discourse.

2.2.2. Genre and SFL

Systemic functional linguists have offered important insights into the generic features of scientific writing (Halliday, 1987; Halliday & Martin, 1993). Halliday (1994) proposed three components of register—field, tenor, and mode—and a genre should be understood from linguistic features associated with these three factors. Field refers to the subject matter or content under discussion (e.g. different academic fields and disciplines); tenor refers to the relationship between text producer and text receiver (e.g. student writing to teacher or expert to expert); mode refers to the channel or medium of communication (e.g. written vs. spoken communication). Naturally occurring texts (both written and spoken) simultaneously entail these three contextual variables. Corresponding to the three situational factors of a register are three metafunctions of language—the ideational, interpersonal, and textual. Thus, a genre linguistically and semiotically represents the three metafunctions as determined by the register, with SFL-informed genre analysis of academic texts mainly focussing on examining linguistic features that contribute to particular metafunctions.

Although genre theory in the SFL tradition has been developed mainly for tackling language education problems at the elementary and secondary level in Australia and recently in the US, analytical frameworks informed by SFL have been increasingly employed to approach academic texts. The SFL approach to genre analysis deals with all three metafunctions represented in academic writing. Although SFL contends that any choice of linguistic device realises all three metafunctions simultaneously, different conceptualisations have been developed to focus on each.

The ideational property of academic writing is examined by looking at transitivity types and participant roles unique to this genre. By studying the evolution of scientific writing and comparing it with everyday speech, SFL scholars have found crucial grammatical devices conducive to meaning making and scientific theorising, of which the use of nominalisation and what Halliday calls grammatical metaphor plays the most instrumental role (Halliday & Martin, 1993). These two devices, along with other information-integrating mechanisms (Chafe, 1982), can account for much of the semantic abstraction and complexity characteristic of academic writing. This will be examined later on in more detail in reviewing SFL's revelations about the complexity of scientific writing.

The interpersonal representation has been mostly approached through the appraisal framework (Martin & White, 2005), an offshoot of SFL specifically elaborating on the interpersonal, evaluative function of language. Recognising that academic writing is not an absolutely impersonal and objectified process has provided impetus for the identification of how research writers embed evaluation and attitude in presenting facts and argument. While expert writers tend to constantly attune themselves to making skilful use of evaluative resources, it is something that could be easily overlooked by L2 and inexperienced writers without being explicitly taught.

Finally, the textual metafunction as embodied in thematic development contributes to text cohesion and coherence. But it needs to be noted that coherence is more difficult to achieve as it involves more than textual smoothness, with the structuring of argument and logic development playing more crucial roles. Consciousness of the systematic arrangement of thematic components across discourse can guide the writer in putting lexicogrammatical resources into appropriate syntactic slots at the clause and sentence level. Features of thematic alternation in academic

discourse have strong implications for the syntactic construction of clause-initial elements, with the use of complex noun phrases being an important device for relating to previous discourse. This ideational feature inevitably adds to text abstraction and authorial objectivity at the interpersonal level. Thus, the three metafunctions forming an interlocking continuum constitute the generic features of academic discourse.

2.3. Research on Dissertation Writing

Thompson (2012) provides a comprehensive review of research into dissertation writing at the master's and doctoral level, although the focus is on doctoral dissertations. The topics covered in Thompson's review include generic, rhetorical, sociocultural, and pedagogical issues concerning dissertations. Although still being an underrepresented research area (Starfield & Ravelli, 2006), dissertation writing has begun to gain more attention as a result of increasing enrolment of international students in research degree programs (Council of Graduate Schools, 2013). This section provides a review of studies on dissertation writing and education, aiming to identify current knowledge of this area and aspects needing careful investigation.

2.3.1. Organisational Structure of Dissertations

Research into the organisational structure of dissertations is informed by the ESP approach to genre analysis, trying to identify variations across disciplines and research paradigms. Paltridge (2002) has identified three major types of dissertation organisation: traditional simple, traditional complex, and topic-based.[8] Table 2.1 presents the organisational pattern for each type based on Thompson (2012, p. 285).

[8] There is actually a fourth type, compilation of research articles. Although it is common in the sciences (Swales, 2004), it is relatively rare in the humanities, hence its exclusion from the review.

TABLE 2.1

Three Types of Organisational Structures of Dissertations

Traditional simple	Traditional complex	Topic-based
Introduction	Introduction	Introduction
(Literature Review)	Literature Review	Chapter 1: Topic 1
Methods	Chapter 1: IMRD	Chapter 2: Topic 2
Results	Chapter 2: IMRD	….
Discussion	….	Conclusion
(Conclusion)	Conclusion	

Note. IMRD = introduction, method(ology), results, and discussion.

The first of these follows the structure of a scientific journal article in the classic IMRD format, with separate chapters for the introduction, methods, results, and discussion; but dissertations are slightly different from journal articles in that they often have a separate chapter for literature review and conclusion. The traditional complex format begins with an introduction and general literature review followed by a number of chapters in the IMRD pattern reporting different studies, then finally finishing with a general concluding chapter. The scale of research projects for which this type of organisational structure is suitable is usually too large to be covered by the traditional simple format. Therefore, the traditional simple is more appropriate for master-level research or moderate-scale doctoral projects, whereas the traditional complex is a better candidate for large-scale doctoral dissertations, however both being designed for positivistic empirical research in the hard sciences and social sciences. By contrast, as identified by Starfield and Ravelli (2006), the topic-based dissertation, beginning with an introduction chapter and then a series of chapters each based on a topic, is common

in the arts and humanities such as music, history, and literature, which do not easily fit into the IMRD paradigm. Dissertations in applied linguistics which extend over the humanities and social sciences can therefore appear in both traditional and topic-based structures, although the former type is more commonly seen.

The differences in organisational structure may have corresponding linguistic concomitants. Dissertations following the traditional IMRD-based patterns are expected to strictly observe a style of "the classic detachment of positivism" (Starfield & Ravelli, 2006, p. 224), in which the researcher stays away from the research process and is discouraged to use the first person pronoun *I* or *WE* in delivering arguments and describing methodology; the researcher is linguistically hidden behind the scenes, leaving the discourse replete with impersonal expressions. Contrastingly, writers of humanities dissertations following the topic-based format may include narrative elements and the writerly self, particularly first person pronouns referring to the research writer who "*interacts* and *intervenes* in his or her research objects" (ibid, p. 230, emphasis original). However, it needs to be noted that despite the emerging presence of the writerly self as shown in the use of first person pronouns, academic writing is still dominated by a highly nominalised grammar and abstract meaning making (Biber & Gray, 2016).

2.3.2. Discourse Features of Dissertations

Discourse features specific to dissertations have also been identified, as evidenced from the lexicogrammatical patterns decided by and contributing to the identity and communicative purposes of the student writer. However, there has not been as much discourse analysis of dissertations as work done for research articles. Swales (1990) points out that discourse analysis of dissertations has been "largely, avoided, at

least partly because of the daunting size of typical text". Since communicative purposes are best distinguished by the different rhetorical sections in the dissertation, linguistic features associated with these variables have been subjected to careful description and evaluation. Comparisons are often made across disciplines and between dissertations and journal articles within the same discipline in order to see how specific meaning-making linguistic devices are used to achieve communicative purposes unique to dissertations in different disciplines. Another important aim of research along these lines is to identify problems in achieving form-function agreement encountered by inexperienced student writers.

A major theme in the slim body of research on dissertations is metadiscourse (Bunton, 1999; Hyland, 2004; Swales, 1990), which serves to play a crucial role in establishing the author's relationship with the evolving text and the target audience (Hyland, 2005). Swales (1990) points out that dissertations, due to their greater length, would use more metadiscourse than research articles as authors need constantly orient their readers to textual sources. Bunton (1999) investigated the use of metadiscourse in 13 EFL PhD dissertations to guide readers and found students tend to use more metadiscourse at the overall dissertation level than at the chapter level, where he suggested student writers to further develop metadiscourse use. Hyland (2004), through examining the functions and distributions of metadiscourse in 240 masters and doctoral dissertations by Hong Kong students, proposed a model of metadiscourse as mainly interpersonal (rather than textual) guides to propositional meaning in different disciplines and genres and suggested ways in which L2 research students' consciousness of metadiscourse use can be enhanced.

Other discourse-oriented research on dissertations examines students' use of functionally meaningful linguistic features for various rhetorical purposes of the

dissertation, such as writers' evaluation and positioning exhibited in literature reviews (Gil-Salom & Soler-Monreal, 2014; Xie, 2016), metadiscourse in results and discussion chapters (Lee & Casal, 2014) and introduction (Kawase, 2015), lexicogrammar used for citations (Charles, 2006; Thompson, 2001; Thompson & Tribble, 2001),[9] purpose statements (Lim, Loi, Hashim, & Liu, 2015), and stance-neutral formulations (Sawaki, 2014).

These studies have revealed useful information on how student writers linguistically achieve certain communicative purposes for the research dissertation—successfully and less successfully, which can be critically taken up for dissertation writing instruction. However, a problem with this line of research is that it is mainly focused on investigating the role of specific linguistic markers in fulfilling local or global discourse functions, without much consideration of the propositional quality or sophistication embedded in the communicative functions concerned. It seems to be premised on the belief that L2 student writers would be adequately equipped for EAP writing through familiarity with learning to use functional signposts in discourse. To confirm this, it would be necessary to understand the kinds of linguistic difficulties and

[9] Coined by Michael Halliday, the term **lexicogrammar** (adj. lexicogrammatical) is a blending of lexis and grammar and used to emphasise the continuity between the two. Different from many linguists and grammarians who treat lexis and grammar as separate, Halliday brings them together. This approach is adopted in this study. To illustrate, the divide between lexical and syntactic categories as traditionally defined would be blurred by the fact that the two sometimes overlap. For example, while lexically derived from verbs and adjectives (and sometimes also from nouns), nominalisations usually function to construe finite clauses as noun phrases, playing a basically grammatically role in this process. By contrast, the use of complement clauses is largely subject to the words they complete, as in the following examples where the use of complement clauses is determined by the **verbs** or **nouns** in front:
(1) *The **problem** is that students may not necessarily realize they are in possession of this potentially very useful L1 meta-awareness for L2 development*.
(2) *Following the **belief** that human brains have the same language faculty for L1 and L2 (perhaps L3 as well), Cummins (1979) proposed….*
(3) *Norton (2000) **contends** that "power relations play a crucial role in social interactions between language learners and target language speakers" (p. 12)*.
It can thus be said that the three words *problem*, *belief*, and *contend* entail a semantically driven grammatical property of taking complement clauses, evidencing the continuity of lexis and grammar. In addition, this term has also been adopted by many corpus linguists.

challenges perceived by L2 students writing dissertations in English. The following section reviews research investigating this issue.

2.3.3. Surveys of Dissertation Writers

A large body of research on dissertation writing has been devoted to surveys on the "agents" of this genre, i.e. student writers, in relation to their needs (Cooley & Lewkowicz, 1997), perceived difficulties (Bitchener & Basturkmen, 2006; Lee & Tajino, 2008; San Miguel & Nelson, 2007), composing processes (Shaw, 1991), writer identity dynamism (Cadman, 1997; Starfield & Ravelli, 2006), and reading efforts (Kwan, 2008, 2009), as they proceed with the dissertation from immaturity to experience. This section will focus on students' perception of difficulties and challenges in writing the dissertation and how well existing pedagogical practices address their concerns.

The issue of students' writing needs is usually studied in association with their perceptions of the difficulties and challenges in the writing process. A common finding is that students takes their challenges and problems more in the use of local lexis and grammar for meaning making and argumentation. In contrast, challenges at the global, macro-structural level are perceived to be smaller (Bitchener & Basturkmen, 2006; Lee & Tajino, 2008; Qian & Krugly-Smolska, 2008; Yeh, 2010). For example, two of the participants in Qian and Krugly-Smolska's (2008) study suggested limited ability to convey information with concise, less verbose use of language. They often had to use a couple of sentences to present a point which they believe more competent writers can do with just one sentence.

The issue here is not one of producing grammatically well-formed sentences as some may simplistically believe, far from it, but how to deploy and coordinate the

lexicogrammatical resources offered by EAP to best present the dissertation writer's meaning potential (Halliday, 1978). To illustrate, many NNS scholars' manuscripts are not linguistically eligible for publication until heavily edited by professional editors (Burrough-Boenisch, 2003). Research methodologies and results being almost intact, it is precisely the reshaping of lexis and grammar that contributes to the publishability of research.

However, students' perceptions often run counter to supervisors' intuitions and hence the focus of pedagogical attention, which tends to attach more importance to teaching macro-level rhetorical structures (Evans & Green, 2007). Bitchener and Basturkmen (2006) revealed that although both supervisors and students considered skilful manipulation of vocabulary and grammar to develop clear ideas and arguments as a major difficulty, their understandings differed: while students saw the difficulty at the sentence level their supervisors regarded the macro-level rhetorical and schematic structures as more challenging. This is not hard to understand since supervisors as experienced research writers themselves are supposed to have passed the phase when they struggled with lexicogrammar during text production; hence they may not vicariously experience the real-time problems students encounter in writing the dissertation on a moment-by-moment, word-by-word basis. Lee and Tajino (2008) interpreted students' perception of language-related difficulty in terms of their "low confidence with the language skills required in English academic writing" (p. 7), which they argued is an internal attributional factor having a negative impact on students' self-efficacy in writing academic English. While low self-efficacy does hinder fluency, it is more important to explore linguistic causes for this affective barrier and to pedagogically prepare students for raised awareness of features of academic English, both at the local and global levels.

2.3.4. Pedagogy and Curriculum for Dissertation Writing

While interest in dissertation writing instruction dates back to the 1980s (e.g. Richards, 1988), efforts to consider pedagogical issues from different angles are relatively recent. Materials development, course design, and writing facilitation needs to be informed by relevant empirical research, particularly studies in relation to student needs and writing experiences, examination of linguistic and rhetorical features of this text type, and knowledge about effective instructional methods. However, Paltridge (2002) found that only a few guides or handbooks had based themselves on empirical scrutiny of existing dissertations and identified a wider range of dissertation types than the books had suggested. In addition, most course design studies have not been broadly based on findings from research investigating the writing challenges of research students as found above, although they involve student writers' introspection and retrospection with the course or writing group designed for them (e.g. Aitchison, 2009; Li & Vandermensbrugghe, 2011; Manalo, 2006).

An exception is Allison, Cooley, Leokowicz, and Nunan (1998), who developed a writing program at the University of Hong Kong based on survey data collected from supervisors and students on their experiences with dissertation writing. Allison et al. identified student problems both at local lexicogrammatical level and at macro organisational level, but local problems, according to the supervisors, did not damage text comprehensibility in any significant way. Here, once again, comes in the issue of discrepant understandings of writing challenges between students and supervisors. In any case, it would be advisable for curriculum and materials developers to synthesise and refer to previous research into needs analysis and student/supervisor perceptions of dissertation writing as important background information on which to ground their work.

Also of note is the fact that most courses and handbooks designed for dissertation writing have taken a genre analysis approach focussing on facilitating the development of move structures, with little attention paid to raising students' awareness of the linguistic features of academic prose. Charles (2007) recommended the combination of top-down, genre-based approaches and bottom-up, corpus-based activities to the pedagogy of postgraduate writing, accommodating the rhetorical functions and their linguistic realisations respectively. The form-focused dimension in Charles' study was mainly based on concordance searches of lexicogrammatical choices, aiming to raise students' awareness of function-form relations. Other corpus-assisted writing programs and materials development also employed context-sensitive corpus searches as linguistic resources and evidence to complement the instruction of rhetorical organisations and move structures (e.g. Chang & Kuo, 2011; Yoon, 2011).

Few studies have discussed ways in which the presentation of complex ideas and argument should be pedagogically delivered to students. It follows that for many educators, course design for teaching dissertation writing does not need to be much concerned about linguistic features other than textually and interpersonally important aspects. The complex task of presenting complex ideas is largely left to students' own devices. But if dissertation writing pedagogy is to be more useful, the kinds of difficulties perceived by research students have to be more directly addressed. Thus, linguistic analysis of the ideational complexity of academic texts, which seems to baffle research students most, needs to be undertaken to inform dissertation writing instruction, as a complement to functionally-oriented revelation of various significant discourse markers reported above.

In summary, dissertation writing has been approached as an independent genre with its own textual and rhetorical characteristics. Studies examining dissertation

writing have mostly focused on its macro-structures and move-step flow in different rhetorical sections, various functionally driven discourse features, difficulties perceived by supervisors and students, and pedagogy and curriculum design for dissertation writing. Surveys conducted with student writers and their supervisors concerning the writing challenges have revealed somewhat different understandings: while supervisors tend to see students' difficulties at the organisational and rhetorical level, students perceive linguistic presentation of complex ideas and argument at the lexicogrammatical level to be more challenging. However, little attention in course and materials development has been paid to address students' concerns about language problems per se, i.e. to express complex ideas and argument in appropriate lexis and syntax (cf. Evans & Green, 2007). If dissertation writing by international students is to be examined within the general domain of second language writing, a good knowledge of its lexicogrammar and how it comes about as a result of combined factors would be a necessary step to better prepare L2 postgraduate students to be more linguistically conscious and effective dissertation writers.

2.4. Syntactic Complexity of Academic/Scientific Writing

As shown in the previous section, research into dissertation writing has predominantly taken an approach that focuses on the examination of its organisational structures and rhetorical features specific to this text type. Contrastive analysis is often adopted to compare the representation of these features in dissertations and research articles or between dissertations of different disciplines. However, in terms of the three components of register proposed by Halliday most research on dissertation writing is devoted to examining the interpersonal and textual aspects of language use, with a neglect of the ideational complexity in novice writers' texts. As mentioned before, this

neglect might be explained by supervisors' and researchers' assumption that advanced second language writing at the postgraduate level is not so much concerned with producing accurate and complex meaning as with attending to the more delicate interpersonal and pragmatic aspects to better achieve communicative goals. However, such an assumption better suits the situation of native English speakers writing in their mother tongue whose major concern about writing improvement is with nuanced qualification of ideational meaning using interpersonal devices. For L2 writers, developing delicate and attitude-loaded meaning making needs to go hand in hand with learning the arrangement of syntactic structures unique to academic writing. Being able to "syntactically resemble" expert academic writers would be the first step to effective academic writing.

This section will first review the grammatical features of academic prose as compatible with its social and textual correlates and then moves to focus on those features that contribute to the ideational complexity of academic texts. The review covers both theoretical and empirical analyses of the grammatical features of academic discourse, both of which have adopted a register variation approach comparing writing with speech. Theoretical observations derive from functional linguistic analyses of the grammatical features of academic discourse (e.g. Chafe, 1982; Halliday, 1989; Halliday & Martin, 1993) and philosophical illustration of human thought modes (Bruner, 1986) while empirical evidence is based on corpus linguistic approaches to the issue (e.g. Biber, 1988; Biber et al, 2011).

2.4.1. What is Syntactic Complexity?

As noted by Biber et al. (2011), the term *complexity* has been widely employed in different subfields of linguistics, but with distinct connotations (see p. 6, footnote 1).

Within applied linguistics, it refers to the variety and sophistication of syntactic structures and lexical items exhibited in written and oral production. Using complex syntax and lexis is one of the goals of language learning and the measurement of syntactic and lexical complexity has long been used to monitor language development (for both L1 and L2). Before reviewing functional and corpus linguists' characterisations of syntactic features of academic writing and the vast array of syntactic complexity measures developed for tracking L2 writing development, it is necessary to be clear about the basic components of syntactic complexity in its general sense.

Two fundamental types of syntactic complexity can be distinguished: phrasal complexity and clausal complexity (Staples, Egbert, Biber, & Gray, 2016). Phrasal complexity is realised by adding modifiers to the head word of a noun phrase, adjective phrase, verb phrase, adverb phrase, or preposition phrase. Of all these phrase types, the possibility for the noun phrase to be significantly expanded is the largest as both premodifiers and postmodifiers can be attached to it.

Clausal complexity, on the other hand, is realised through clause combination, by means of hypotaxis (i.e. adverbial clauses and non-restrictive relative clauses; see Ex. 3 and 4 in this chapter) rather than embedding (complement/nominal clauses and restrictive relative clauses; see Ex. 6, 7, and 5). However, there is a borderline situation for embedded complement clauses (e.g. *that*-complement clauses). Unlike adverbial clauses that enter into logical relationships with their matrix clauses (e.g. reason, condition, and concession), complement clauses do not form such relationships with the superordinate clauses of which they are a part. Instead, they are used as clause elements: 1) as a subject (Ex. 6), 2) embedded in the predicate and controlled by a verb, adjective, or preposition (Ex. 7), and 3) embedded in a noun phrase to expand the

meaning of the head noun (Ex. 5). Therefore, complement clauses can either be embedded in a noun phrase (as in Ex. 5) or stand on its own as a nominal structure (as in Ex. 6 and 7). Now that syntactic complexity has been viewed from two distinct perspectives, the review of functional and corpus linguists' claims on the grammatical features of academic writing will also be examined in relation to this distinction.

2.4.2. Linguistic Analyses of Written Syntactic Complexity

Different from formal theories of linguistics that take grammar mainly at the structural level, functional linguists see grammar as an inventory of meaning-making resources. Therefore, a particular genre would select resources available in the grammar of a language in ways that are conventionally valued in the genre's discourse community and it is the responsibility of the writer/speaker to create an expected style by making appropriate choices. Although one may tell that academic writing is characterised by the use of technical lexis and complex syntax demonstrating a style of formality and abstractness, it would be more important to see what specific grammatical structures and syntactic arrangements result in such features and how the choice of linguistic resources relates to its contextual and functional demands.

This section reviews the work of functional linguists and discourse analysts that explores the grammatical features of writing as compared with those of speech (Chafe, 1982; Chafe & Danielewicz, 1987; Halliday, 1987; Halliday & Martin, 1993). At the same time, their insights are constantly related to evidence from corpus-based investigations into the grammatical features of academic writing, notably by Biber and his colleagues (Biber, 1988; Biber et al., 1999; Biber et al., 2011).

2.4.2.1. Cognitive and Functional Perspectives: Some Basic Common Ground

Linguists examining features of writing from cognitive and functional angles have observed that academic writing, compared with informal speech, is informationally integrated and condensed, both in terms of noun phrase complexity and clausal complexity. This section reviews the lexicogrammatical devices functioning to condense and integrate information and their cognitive and metafunctional foundations. The review starts with the work of American cognitive linguist Wallace Chafe and proceeds to that of systemic functional linguists.

Proceeding from the observation that writing allows for time to prepare for integrated expression while speaking does not, Chafe (1982, pp. 39-44) and Chafe and Danielewicz (1987, pp. 94-105) identified a number of lexicogrammatical features of writing that contribute to the integration of ideas. [10] These are the use of nominalisations, participles, attributive adjectives, conjoined (coordinate) phrases, series (coordination of more than two phrases), sequences of prepositional phrases, complement clauses, relative clauses, passives, and abstract subjects. It can be seen that most of the linguistic devices identified by Chafe serve to expand the complexity of the clause rather than to enact clause combination. More specifically, it is primarily the complexity of the noun phrase rather than other phrase types that is augmented by using these devices, since NPs constitute the major ideational components of a clause. The following examples illustrate how the use of these devices expand the clause via noun phrase complexity:

(8) *The* ***further***[attributive adjective] ***analysis***[nominalisation] [*of such dominant semantic categories in context through concgrams and concordances*][sequence of prepositional phrases] *provides a window into the* ***underlying***[participle] *ideological assumptions and*

[10] Although Chafe intended to make claims for the written registers in general, the data he selected for investigation was academic writing.

values of the linguistic manifestation, and [*whether such assumptions and values have changed over time*]$^{\text{complement clause}}$.

Although Chafe recognised complement clauses as characteristic of written registers, which, as previously discussed, is on the borderline of clausal complexity, the use of hypotactic subordinate clauses, adverbial clauses in particular, did not appear in the list. He did not explain why such subordination was not included as a written feature, neither did he list that as a feature of oral interaction. However, as will be shown later, research on the syntactic complexity in student writing has attached great importance to the complexity of subordination. Despite a preliminary study based on a small dataset, Chafe's observation of the linguistic features of academic writing was keen and systematic, laying a sound foundation for subsequent research.

Another line of functionally-directed research into the syntactic complexity of writing is seen among the work of systemic functional linguists, who are particularly interested in revealing the variability of register configurations among spoken and written modes of discourse (Halliday, 1987, 1989; Halliday & Martin, 1993; Halliday & Matthiessen, 1999). Central to this body of research is the view that both speech and writing are complex, but in radically different ways. The complexity of speech has been characterised as "intricate", primarily through the frequent use of subordinate and coordinate clauses (see the speech extracts and analyses in Halliday, 1987, 1989), while writing is complex in lexical density rather than grammatical intricacy. The grammatical structure of writing is in fact quite simple, with many sentences being independent clauses and containing only one main verb, for instance:

(9) *This article* ***examines***$^{\text{main verb}}$ *the multilingual and multimodal practices of British Chinese children in complementary school classes from a multicompetence perspective.*

Lexical density is a measure of the number of lexical words (i.e. content words as opposed to grammatical/function words) as a proportion to the number of clauses in a discourse,[11] and by "grammatical intricacy" is meant the use of grammatical words and the incorporation of dependent clauses. Halliday (1987) found that in expressing a fixed set of meanings, whether it be narration or exposition, lexical density increases as the text moves towards higher formality (from speech, through personal letters, to academic writing). This is so because formal writing uses much fewer clauses and grammatical words than spontaneous speech, hence more lexical words per clause.

Speech, in contrast, involves a stack of coordinate and subordinate clauses at the sentence level, thus bringing down its lexical density and increasing its grammatical intricacy. It is also noteworthy that it is the parsimony in using clauses at the sentence level that characterises written registers, as shown in Ex. 9 above. Lexical density and grammatical intricacy can be imagined as moving towards opposite directions on a formality continuum, with the most informal genre of spontaneous speech and the most formal genre of writing at the two extremes and many intermediate genres in between (Tannen, 1982).

Similar to Chafe's (1982) observation, Halliday found that the key to the heightened lexical density of writing is the prevalent use of nominalisation and pre- and post-modification to form complex noun phrases (CNPs).[12] In fact, a number of the linguistic devices identified by Chafe as characteristic of academic writing function as modifiers to complexify the noun phrase, i.e. attributive adjectives and participles as premodifiers and postmodifiers, series of preposition phrases, (noun) complement

[11] Although lexical density can also be calculated as a proportion of the number of all running words in a discourse, the contrast between speech and writing is much more prominent with the clause-based measure.

[12] In SFL, noun phrases as used in traditional grammar are referred to as nominal groups (see Halliday & Matthiessen, 2004 for the difference between group and phrase). The present study sticks to the term noun phrase in this regard.

clauses, and relative clauses as postmodifiers (see Ex. 8 above). Thus, what is expressed in clauses in speech would be expressed in noun phrases in writing and the use of CNPs helps expand the clause and reduce the number of clauses needed. For example, the clause 10a can be condensed into a noun phrase (10b):

(10a) The students used "come back" instead of "go back"

(10b) the students' use of "come back" instead of "go back".

However, with clauses compressed into noun phrases, also disappearing are the logical relationships between clauses, especially those between dependent clause and matrix clause. These relationships may include such conceptions as time, reason, causality, and location. Taking a step further than Chafe, Halliday pointed to a crucial mechanism that recodes these relationships and "gives life" to nominalisation and complex noun phrases in their realisation, that is, *grammatical metaphor*, already mentioned in 2.2. This will be introduced in detail below.

Simply put, grammatical metaphor refers to using language in such a way that processes and states of affairs originally encoded in verbs, adjectives, and prepositions are nominalised, *as if* they become things and concepts that can be further predicated (such as the noun phrase in Ex. 10b above). Nominalised clause participants, mostly complex noun phrases, become abstract and static entities, as a result of the removal of agents, tense/aspect, and modality, and of the semantic indeterminacy between nominalisations and their modifiers.

The level of abstraction is further enhanced when the logical relationships between clauses are recoded as verbal processes. For example, *they are rich so they can stay in a luxury hotel*, via grammatical metaphor, becomes *their wealth allows them to stay in a luxury hotel*. This simple example illustrates how grammatical metaphor works for written language, and anyone found constantly speaking this way in everyday

situations would be thought of as deviant. While the message remains much the same, the wording and the mode of meaning making do not. Scholarly academic writing in any discipline is replete with grammatical metaphor, which is seen as the single most distinctive characteristic of academic writing relative to speech. The underlining and bold type in the following extract from a journal article in applied linguistics serve to illustrate, where the underline marks complex noun phrases and the bold type highlights verbal processes:

(11) *Analysis of verbalizations* ***originated*** *as an investigative method for information processing (IP) approaches to psychological research. SLA research taking an IP approach* ***investigates*** *learner verbalizations as a means to gain access to learners' cognitive processes. Findings of nearly two decades of research in this tradition* ***have elucidated*** *relationships between attention and awareness and SLA.*

Halliday has employed a number of dichotomies to describe this difference. He characterised speech which is rich in hypotaxis and parataxis as "process" and writing which is full of abstract relations as "product" (Halliday, 1987, p.74). Described in adjectives, speech is "dynamic" and "congruent", writing is "synoptic" and "incongruent". By "congruent" and "incongruent" Halliday emphasises that processes and characterisations expressed through clauses harmonise with the outside world, whereas those coded as nouns do not. It is this kind of incongruence with human beings' ordinary perspectives of the world that occasions the complexity of writing.

Thus, it might be more appropriate to understand syntactic complexity from a cognitive rather than from a grammatical point of view, with complexity synonymous with being difficult to produce and comprehend. Although Halliday has revealed that spoken language is more complex than writing in using grammatical words and

dependent clauses, the level of cognitive sophistication involved in producing and comprehending written texts caused by grammatical metaphor is considerably greater. The "intricacy" property typical of spoken language could just have arisen out of the uncontrollable verbosity and unplannedness of spontaneous production.

Interestingly, the linguistic account of differences between speech and writing as the one provided by Halliday can find similar arguments in cognitive psychology. As noted by Vande Kopple (1994), cognitive psychologist Jerome Bruner (1986) proposed two modes of knowing reality: the narrative mode and the paradigmatic or logico-scientific mode. Each mode corresponds to one of two spheres of reality: "that of nature and that of human affairs" (ibid, p. 88). The narrative mode of thought corresponds to the sphere of human affairs while the paradigmatic mode of thought to the sphere of nature. Without much penetration into Bruner's theory, it suffices to note that many characteristics of human cognition in the paradigmatic mode open up the possibility of enacting grammatical metaphor. But of course they also explain other higher-order human cognitive behaviours such as categorisation and mathematical thinking. However, although unique to humans, the paradigmatic mode of thought or the metaphorical mode of writing does not take shape naturally but needs explicit learning of discourse representing this mode. Explicit instruction seems particularly important for students from working class backgrounds who are often deprived of the written code due to their underexposure to it, according to Bernstein's code theory in the sociology of education (Maton & Muller, 2007).

In an L1 context, regular exposure to formal language use starts from elementary school, where the pedagogy of basic content subjects relies largely on textbook input, facilitated by teachers' explanation and demonstration. This is when

children have to grapple with a whole new set of language use, the language of schooling (Schleppegrell, 2004). As Halliday (1993) noted:

> When children learn to read and write, they have to enter a new phase in their language development. In the process of becoming literate, they learn to reconstitute language itself into a new, more abstract mode. Reconstituting language means reconstituting reality: Children have to reinterpret their experience in the new mode of written language. This is not just a matter of mastering a new medium . . . it is mastering a new form of knowledge: written educational knowledge as against the spoken knowledge of common sense. (p. 109)

It is approximately since junior-secondary school that knowledge begins to be conveyed through discipline-based coursework, broadly divided into the sciences and humanities (Halliday & Martin, 1993). By this time, students need to process, to a lesser extent produce language that is used to identify, exemplify, and classify—functions to construe the intellectual world. As subject knowledge advances over the school years, the language representing this knowledge proportionately grows more complicated in both lexis and grammar. Schleppegrell (2004) has shown that even texts used for elementary and secondary schooling pose great challenges for children if they are not offered opportunities to be acquainted with the linguistic features of such texts and their corresponding functions. Therefore, it goes without saying that the academic code presents a much greater challenge for non-native speakers for whom exposure to academic discourse would not begin till university.

2.4.2.2. Metafunctions of Grammatical Metaphor

The previous section has outlined some of the most distinctive linguistic features of academic writing identified by cognitive and functional linguists interested in register variation. It has been seen that linguistic devices used to form the noun phrase occur frequently in academic texts and the noun phrases are usually highly complex with pre- and post-modifiers, i.e. complex noun phrases (CNPs). It has also been illustrated that CNPs typical of academic writing are actually a contraction of verb-based clauses typical of spontaneous speech. The mechanism that facilitates this process is called by Halliday grammatical metaphor, in which nominalisation and the figurative use of verbs play a crucial role. This section continues the discussion by focusing on three metafunctions (i.e. ideational, interpersonal, and textual) of grammatical metaphor to illustrate its legitimacy in academic writing. The presentation is based on Schleppegrell (2001, 2004) who, using SFL as an analytical tool, examined the linguistic features of the language of schooling as contrasted with informal interaction, aiming to reveal the challenges of school-based texts for young students who are unfamiliar with the features of this language. Differences were found at the three meaning-making levels between the two registers.

At the ideational level, besides the fact that school-based texts use more technical lexis, more interesting is the finding concerning the ways academic registers condense complex information and have it structured in logical relationships. Academic language is characterised by dense phrasal structures that pack information and arguments in highly integrated ways. Unlike conversation where logical connections are usually made through conjunctions (e.g. *and*, *but*, *so*, *because*, and *if*.) to establish coordination and subordination, school texts use more nominalisations in subject and object positions connected by verbs. In English, nominalisation is the process by which

verbs and adjectives are transformed into nouns through derivational suffixation or conversion (i.e. zero derivation), so it is capable of transforming clauses into phrases. As such, the information that is presented in conversation using between-clause connections would be restructured in academic texts in within-clause manners. This will pack more information into each clause, allowing the text to expand efficiently.

At the interpersonal level, the use of nominalisations and metaphorised verbal processes presents a style that is objective, authoritative, and detached. Academic registers do not have recourse to explicit prosody and emotion available in spoken interaction for the conveyance of attitudes and stances of the author. However, this is not to suggest that academic writers take a purely non-interacting and introvert stance towards the text and the imagined audience. Rather, stance taking in academic texts is achieved by careful selection of lexicogrammatical resources, along with typographical devices such as bold and italic typefaces. The selection of resources for academic texts apparently needs to project the author in an authoritative position as a reliable sender of knowledge and information. A straightforward resource for this function is to use formal lexis and syntax, as formality can serve to generate authority, and even awe (Fairclough, 1989). Formality to a great extent depends on using technical terms, and as Halliday and Martin (1993) argue, technicalised discourse is also a grammatical process, in which grammatical metaphor plays a large part. A nominalised grammar turns arguments and concepts into objective "things" to pack not just information but also the author's attitudes and stances implicitly connoted in the noun phrase (Martin & White, 2005). Packing attitudes in noun phrases may sometimes efface the negotiability of claims, by backgrounding certain elements as given facts. For example, instead of saying *his argument is untenable*, one can say *his untenable argument*, thus

the use of the attributive adjective *untenable* to a greater degree perpetuating the author's attitude.

At the textual level, grammatical metaphor is conducive to the creation of text cohesion and coherence. It is useful to refer to the concept of theme-rheme development in the clause to understand discourse flow. Following the Prague school of linguistics, Halliday and Matthiessen (2004) define theme as "the element which serves as the point of departure of the message; it is that which locates and orients the clause within its context", and rheme as "[t]he remainder of the message, the part in which the [t]heme is developed" (p. 64). Theme usually includes known information and rheme develops new information added to the theme. As described by Halliday (1994), the theme of a clause comprises everything up to and including the initial ideational element and the rest of the clause is the rheme. In the following examples, the themes are marked in bold:

(12) ***This excerpt*** *is drawn from a session…*

(13) ***There are two features*** *in the tutor's talk…*

(14) ***By and large, students' bids*** *for assistance are treated unequivocally by tutors….*

(15) ***How does the Chinese MA student's use of lexical phrases*** *develop over an academic year?*

The theme-rheme dichotomy may in some way correspond to the distinctions of topic-comment and given/old information-new information in other semantic and grammatical frameworks, but it is more motivated by a discourse development perspective. In natural discourse, the rheme of the previous message would possibly become the theme of the next for the discourse to be coherently expanded.

Academic writers may employ theme to summarise information in previous discourse. In many cases, this previous information occurs as a clause or sentence and is compressed into a phrase in the thematic summary. Vande Kopple (1994) has found that academic writing frequently uses lengthy CNPs as grammatical subjects through which the author establishes some common ground with the reader, by bringing up given information or shared disciplinary knowledge. The effect of using complex phrasal structures in the thematic portion on overall text coherence and quality is supported by Crossley, Roscoe, and McNamara (2014) who found that essays (by English L1 students) with long NP subjects tended to be rated higher for coherence by human raters. In the next section, the use of long subjects in academic discourse will be examined in more detail.

To summarise, the role of grammatical metaphor in academic writing can be looked at from the perspective of its ideational, interpersonal, and textual metafunctions. In academic writing, the three metafunctions, as in any naturally occurring discourse, take effect simultaneously. Academic discourse today, whether in the humanities or the sciences, has gone through an evolution of around four centuries since the scientific revolution in Europe (Halliday & Martin, 1993). In this process, a major innovation in the development of a discourse of science has been with the technicalisation of lexis and grammar, during which emerged grammatical metaphor. Halliday (1993) writes that "…grammatical metaphor is the key for entering into the next level, that of secondary education, and of knowledge that is discipline-based and technical" (p. 111). A nominalised grammar condenses information and is able to link together chains of reasoning that is distilled into nominals. The choice of such lexicogrammatical resources, as described above, has significance for all the three metafunctions congenial

to academic discourse and this has distinguished formal writing from everyday interaction.

While the critical angle of SFL sees this highly technicalised grammar as "engendering only prestige and bureaucratic power…[and] privileging the expert and limiting access to specialized domains of cultural experience" (Halliday & Martin, 1993, p. 17), in terms of being apprenticed into the academic register it would still be a necessary step for novice writers to have explicit awareness of its linguistic features. Being able to use technicalised lexis and grammar is not just a matter of tradition, because as Halliday and Martin argue, a technical grammar also serves "as a means of arriving at new knowledge, a resource for enquiring and for thinking with" and "is actively engaged in bringing such [pre-existing conceptual] structures into being" (ibid, pp. 6–9). The language of science is an instrument not just for reflecting scientific experience but also adds new meanings to it.

2.4.2.3. Inanimate and Abstract Subjects in Academic Discourse

One of the syntactic features of academic English, as contrasted with everyday speech, is the use of inanimate/abstract subjects realised by nominal structures (e.g. noun phrases and nominal clauses), such as:

(16) [*This improved mastery*]$^{\text{inanimate subject}}$ *entails a number of elements*….

Biber et al. (1999, p. 378) show that over 60% of abstract verbs (causative, occurrence, and existence verbs), over 30% of activity verbs, and even communication and mental verbs (20% and 10% respectively) are used with inanimate subjects in academic prose; they also find that inanimate subjects are more frequently used with active verbs than passively, performing two main discursive functions—to show causality and to explain (Master, 1991). This syntactic device constitutes one of the major means for making

the academic genre less agent-oriented—together with the passive construction among others—and its employment has been found to differentiate academic texts produced by L2 users of English from those by L1 students (Callies, 2013). However, as MacDonald (1992) speculates, the use of inanimate subjects (equivalent to what she termed "epistemic" subjects) may even not be a regular part of L1 English students' syntactic-textual repertoire "well into the undergraduate years" (p. 538).

From a functional perspective, there are both discourse-pragmatic and textual concerns that motivate academic writers to use inanimate/abstract NP-subjects. The use of inanimate subjects to background the agentive participants in the discourse can be understood as a prominent representation of grammatical metaphor. As already discussed, grammatical metaphor, realised mainly through the nominalisation of verbal processes, provides a venue where the congruent agent-process semiosis common in personal narratives moves toward a non-congruent, metaphorical pattern. Thus, verbal processes become concepts or things and the human agents involved in the processes fade into the background. Such a change in meaning-making mode allows for the co-existence of multiple nominalised processes in a single clause, whose relationships to each other also serve to contribute to the complexity of propositions thus made.

But it should be noted that the deployment of inanimate/abstract subjects may differ across disciplines. In her study of epistemological differences among three disciplines—psychology as representative of the social sciences and history and literature the humanities, MacDonald (1992) revealed that academic writing in psychology used as subjects more nouns denoting abstraction related to research and reasoning while history and literature (literature in particular) employed more nouns emphasising particularism concerning actual participants under scrutiny. Thus, the implication for the present study is that academic texts in applied linguistics may as

well resemble those in psychology with regard to the use of abstract NP-subjects due to similar research reporting paradigms, i.e. the so-called IMRD pattern, notwithstanding a complex variability of this pattern in actual practice (e.g. Lin & Evans, 2012).

In a study specifically examining the use of particularly lengthy NP-subjects, Vande Kopple (1994) identified three functions of their presence in scientific discourse (his data all coming from the natural sciences though), which he believes to be derived from three sources of pressure on research scientists producing journal articles: the pressure to be precise, the pressure to be concise and economical with words, and the pressure to be efficient and progressive in constructing claims; the last pressure is actually derived from the first two and from the broader socio-semiotic situation in which writing is situated. One might argue that lengthy constituents should not appear in the beginning position of the clause but rather should be placed in the predicate according to the end-weight principle of clause structuring (Quirk et al., 1985). Moreover, what appears later in the clause is usually newer and more important information, which deserves extended sequences of predication (as opposed to given or older information usually appearing at the start of the clause and realised by shorter constituents often involving demonstratives or pronouns), in conformity to the end-focus principle of information processing.

Thus, it may seem that using lengthy CNGs as clause subjects represents a violation of the two principles and should be discouraged in academic writing. However, the two principles should not be considered in the absolute and there is much variation to the configuration of given and new information in the flow of discourse. As Quirk et al. explain,

> In certain circumstances, it is quite normal to have the focus on a noun phrase as subject of a clause, in violation of the end-focus principle. This is frequently because, with the subject concerned, the predicate is relatively predictable and thus has lower communicative dynamism. More broadly, it reflects the fact that nouns generally convey more information than verbs. It is significant that the phenomenon in question is especially associated with intransitives, where (if English structure permitted it freely) we might expect the element order VS [verb-subject]. (p. 1366)

The predictability of the predicate is often grounded in the context and facilitated by conventionalised schemata for understanding discourse. For example, the meaning of the passive predicate in the main clause in Ex. 17 below is easily predictable from the foregrounded concessive clause led by *although*:

> (17) *Although these two studies contribute to the understanding of task planning and attention to form during task-based learner-learner interaction, what learners really do during pretask planning time* ***was not examined***.

Thus, apart from intransitive verbs that may render the predicate more predictable, using the passive voice as in Ex. 17 may also serve to place focus on the subject.

Furthermore, the case of NP-subjects becomes more complicated with respect to the dichotomy between given and new information, because as illustrated by Vande Kopple (1994), while postmodifiers of the head noun may bring up previously mentioned entities or processes, premodifiers usually add new information to them. The example Vande Kopple gave for this point is *additional treatments of this problem*, where *this problem* obviously carries the given information with the use of the demonstrative *this*, but *additional treatments* would bring something new to the problem concerned and hence becomes the focus of the noun phrase. Thus, what

appears in the subject position does not necessarily carry old information only, and even the old information involved in the head noun with its multiple postmodification in the case of much extended noun phrases would fall into a newer category by dint of the premodifier. It can be seen that the three sources of pressure prompting research scientists to produce aberrantly lengthy NP subjects provide convincing evidence of circumstances under which the basic information-structuring strategies based on end-weight and end-focus principles are qualified.

It also needs to be noted that the cognitive demand for producing CNPs subjects would be greater than their instantiations as objects or other complementary constituents in the predicate. This is because the production of objects/complements may be activated or primed by the semantics of subjects and/or verbal elements while the decision on subjects is a relatively more challenging task, as can be seen from children's habitual omission of subjects due to their limited processing abilities during the early stages of first language acquisition (Valian, 1991).

2.4.2.4. Subordination in Speech and Writing

By now it seems that noun phrase complexity constitutes the major source of syntactic complexity in academic writing, but it begs the question if there is any significant role for clausal structures to play. Apparently this has not gained much ground in SFL's expounding of academic discourse, but it deserves more careful consideration for academic writing to be fully characterised. The issue is mainly concerned with whether hypotaxis—identified as a distinctive feature in spontaneous speech—has any contribution to make in academic writing. Before approaching the role of hypotaxis in academic writing, the clausal features identified by Chafe (1982) will be revisited.

Chafe (1982) identified two clause-combining devices apart from phrasal features in his characterisation of academic writing as being integrated: participles and (finite) complement clauses.[13] The English participle is a highly versatile feature, having both phrasal and clausal functions. Apart from its function to premodify (used independently as an attributive adjective, e.g. *conflicting* in *the* ***conflicting*** *outcomes of studies comparing deductive and inductive L2 instruction*) and postmodify (appearing in a participle clause as a non-finite restrictive relative clause, e.g. the clause introduced by *comparing* in *the conflicting outcomes of studies* ***comparing deductive and inductive L2 instruction***) nouns, it can be used in non-finite adverbial clauses and non-finite complement clauses. The former relation is one of hypotaxis (Ex. 18) and the latter embedding (Ex. 19), respectively exemplified below:

(18) *As he concluded with an informal, but accurate, depiction of inchoativity,* ***noting that the windows did not open of their own volition***[non-finite adverbial clause], *the subsequent laughter and requests for student talk and dumbing it down in (l) and (o) suggest that his analysis was inaccessible to many, perhaps even the teacher*.

(19) ***Comparing Figures 2 and 3 furthermore***[non-finite complement clause] *shows similar proportions of analytic talk for individuals in both small-group and whole class Co-construction.*

Chafe's examples of participles (pp. 40-41) from his written data include all of the above circumstances.

[13] Restrictive relative clauses are not considered as clause combination since they are concerned with noun phrases complexity rather than clausal complexity.

On the other hand, finite complement clauses in academic writing most notably appear after so-called reporting verbs, i.e. communication and mental verbs that can take complement clauses, such as:

(20) *Nonetheless, Kirschner et al. (2006) argue* ***that unguided induction is counter-productive if new information overwhelms learners' short-term working memory***[finite complement clause].

However, although clausal features are included in Chafe's analysis, hypotaxis has taken on only a minimal presence in the use of participle clauses as adjunct structures. This hypotactic feature is so rare in speech that it is often studied in isolation from its canonical form, i.e. finite adverbial clauses coding various semantic relations, including time, manner, reason, concession, and condition (Biber et al, 1999, p. 842). As will be shown later, most finite adverbial clauses of these relations are more frequent in conversation than in academic writing.

However, frequency may only indicate ways in which discourse is structured in a particular genre but does not tell much about the nature of the meaning-making process. Although it is true that academic writers use much fewer adverbial clauses, questions like why they still use them at all and whether they use them differently from speakers are not known from frequency counts. Subsequently, differences in using adverbial clauses between speech and writing will be reviewed and a case will be made for including features of subordination specific to academic writing as an integral part of its syntactic complexity; both finite and participle adverbial clauses will be discussed. The review of finite adverbial clauses is based on Schleppegrell's (1996) examination of the similarities between ESL writers' use of subordination and that used in informal speech, while that of participle clauses on Biber et al. (1999) and Kortmann (1991).

Schleppegrell (1996) argued that the use of conjunctions to introduce adverbial clauses is different in speech and writing. Although it has been recognised that spontaneous speech is characterised by extended series of parataxis and hypotaxis while writing by information condensation in the noun phrase, the discourse representations of clause combination in the two modes differ. It would be wrong to assume that academic writers use adverbial clauses in speech-like ways.

As described by Schleppegrell, subordinate conjunctions in spoken discourse usually have "functions related to interactional concerns such as negotiating meaning and information flow, or structuring discourse", and as cohesive discourse markers they "make pragmatic contributions to the ongoing interaction and structuring of discourse, and their semantic contributions are less in focus" (ibid, p. 272). Examining the use of *because* as a conjunction in speech and writing, Schleppegrell identified three uses of *because*-led adverbial clauses that are common in spoken English but seldom seen in writing: "to provide knowledge-based linking, to add information in independent segments, and to link larger segments of discourse" (p. 280). As evidence of the latter two uses, it would often be the case that speakers use *because* to start an utterance that is not a subordinate clause of the previous utterance but has a causal link to a broader segment (or even segments) of previous discourse (see the examples in ibid, pp. 277-279). Thus, they share similar functions with sentence adverbs (e.g. *however*) and the "semantic force" between the discourse segments they connect is usually weak. Frequent use of subordinate conjunctions reflects the speaker's somewhat automatic monitoring of the immediate discourse needs in dynamic oral production, meanwhile adding to the prosodic features of discourse to guide the listener.

By contrast, subordinate conjunctions in academic writing are used to mark meaning relationships between ideationally complex propositions, with pragmatic

discourse-marking functions typical in speech seldom occurring. For the writer, ideational complexity would be greatly enhanced by the need to cater for the proposed meaning relationships between clauses, such as reason, condition, and concession. These relations are crucial to argument building in academic writing as writers need to constantly provide reasons, justifications, assumptions, and defensive stances for certain claims or research actions (Crammond, 1998). The purpose of adverbial clauses in academic writing is to enact these argumentative processes in a textually explicit way.

In addition, adverbial clauses in academic writing would not appear in short consecutive chains as in spontaneous speech, but rather would be expected to encode ample and cogent argumentation to support claims made in the matrix clause. However, providing such argumentation would be a highly challenging task as it requires considerable insider knowledge about the subject matter being written about and strong critical thinking skills. Since it has been found that the quality of critical thinking and therefore argumentation is significantly influenced by content familiarity (Stapleton, 2001), the ability to use adverbial clauses in academic writing to facilitate argumentation reflects not so much syntactic complexity as the sophistication of discourse-level meaning structuring (Crammond, 1998). Although sentence adverbs (e.g. *however* and *therefore*) have similar functions in this regard, they do not link clause-based propositions but rather interact with segments of previous discourse. As such, sentence adverbs would not produce the kind of semantic complexity comparable with that created by adverbial clauses.

In addition, as noted by Chafe (1982), there is a clause combining device that is unique to writing but sounds unnatural in speech: participle adverbial clauses (see Ex. 18 above). Although more typical of narratives (e.g. fiction and journalism), the use of

participle clauses as adjunct is often seen in academic texts as well. This clause type prototypically derives from a finite adverbial clause, a non-restrictive relative clause, or a coordinate clause, with such elements as agent (logical subject), tense, and modality removed. However, the agent is often traceable as it is supposed to be identical with that of the matrix clause. When it is not, there is the danger of forming a so-called dangling clause, deemed ungrammatical in traditional grammar.[14] This feature marks a major difference from its finite counterpart, whose subject can be different from that of the matrix clause. Moreover, this clause type can be further reduced into verbless adjuncts when it describes an attribute of the matrix clause subject, appearing as a noun phrase or adjective phrase attached to the matrix clause, with or without a conjunction:

(21) *It might be inferred that she, as a female L2 learner,* ***although a highly proficient one****, is using formal language forms….*

(22) *Marginal results,* ***although helpful****, are not always clearly interpretable.*

Another major difference from its finite prototype is that the semantic relationship between the participle clause and the matrix clause is often indeterminate, only partially discoverable from the surrounding context. However, this semantic indeterminacy has offered academic writers more flexibility to combine clauses that are related but not in a close manner.

Therefore, due to differences in discourse function and semantic contribution of finite and participle adverbial clauses in speech and writing, it would be advisable to see clause combination in academic writing as a component of its syntactic complexity. In this study, a case is made for treating noun phrase complexity realised by grammatical metaphor and clausal complexity realised by clause combination as

[14] But there are certain cases where dangling can be accepted, especially with the matrix clause introduced by a dummy subject such as *there* or *it*.

different dimensions of syntactic complexity. Although competent academic writers are able to compress much information and logical relationship into noun phrases that is otherwise coded through short clauses in speech, clause combination is still necessary for argument building by explicitly marking justification, counterargument, and concession. The whole purpose of using adverbial clauses is to coordinate logical and semantic relationships between complex statements, signposting important junctions of argument development.

Thus far, the syntactic complexity features of academic writing as described by functionally oriented linguists and discourse analysts have been presented, with devices contributing to noun phrase complexity and certain writing-specific forms of clause combination highlighted. Taking a more empirically driven approach, corpus linguists interested in register variation have also identified high-frequency grammatical features of academic writing, usually achieved by comparison with speech. Subsequently, these high-frequency features will be presented along with their frequency information, which can then be used as quantitative assessment of the observations of functional linguists. Particular attention will be paid to what specific phrasal and clausal features have been identified, and the significance of *frequency* for determining generic features will be briefly evaluated.

2.4.2.5. Evidence from Corpus Linguistics

This section reviews findings of corpus-based studies investigating distinctive linguistic features of academic discourse. Before reviewing some major studies in this area, a brief note is in order on the rationale of corpus-based contrastive studies.[15]

[15] For the differences between corpus-based and corpus-driven studies in corpus linguistics, see McEnery and Hardie (2012, Ch. 1).

Corpus linguistics uses computational methods to analyse sizable authentic data of spoken and written texts to explore the frequency of particular linguistic features and their co-occurrence patterns. Commonly used for corpus-based studies is the contrastive (or comparative) approach to linguistic variation across time or contextual factors, such as different dialects, text types, and language users. This approach is based on the identification of linguistic features that are significantly more (or less) frequently used in one corpus (representing one contextual factor) than in another (representing another contextual factor). Corpora used for this purpose are usually tagged for part of speech (POS) and sometimes parsed as well. For corpus linguists, *frequency* is probably the most important criterion to determine the value of a particular linguistic feature. In fact, linguists engaged in variation studies before the corpus age relied mainly on intuitive perception and observation of the relative prominence of particular linguistic features in different text types, as seen in the work of functional linguists reviewed earlier. Although this intuition should also be a product of psycholinguistically-based frequency effects of exposure to language use, it is the merit of corpus-based variation studies to accurately reveal the frequency profile of a text type.

The syntactic complexity of academic discourse has been closely studied by corpus linguists using the corpus-based contrastive approach. This line of research is mainly taken up by Douglas Biber and his colleagues, who have been taking a sustained interest in identifying linguistic variation across different registers and genres (Biber, 1988, 2006; Biber et al, 1999, 2011). The rationale is that certain linguistic features are typical of academic writing if they are significantly more frequently used than in other registers, notably informal conversation.[16] Biber (1988) identified a set of key linguistic

[16] However, how the frequency criterion can be equated to syntactic complexity needs to be called into question, which will be discussed later.

features in academic writing with his multidimensional approach to corpus analysis. In addition, the *Longman Grammar of Spoken and Written English* (Biber et al. 1999) provides frequency information on every lexicogrammatical feature across four registers: academic prose, fiction, news reports, and conversation.[17]

However, Biber's more focused interest in the syntactic complexity of academic writing is motivated by the belief that research on the syntactic complexity of second language writing has inappropriately employed measures typically concerned with speech rather than writing (Biber & Gray, 2010; Biber et al., 2011). As will be seen later, these are mostly measures of clause combination. Thus, Biber and Gray (2010) challenged two stereotypes of characterising the syntactic complexity of academic writing commonly seen in writing research: 1) that academic writing is "elaborated" in using dependent and embedded clauses, and 2) that academic writing is semantically explicit. To challenge the first stereotype, they used corpus evidence to show that complement clauses (both finite and non-finite) and finite adverbial clauses are much less frequently used in academic writing than in conversation; however, both finite and non-finite relative clauses are much more frequent in writing. They went on to show that in fact relative clauses join "dependent phrase types" to be high-frequency structures in academic discourse, most of which contribute to the expansion of the noun phrase.

The second stereotype was challenged by showing that the use of phrasal structures results in implicitness, rather than explicitness, of the meaning relations among grammatical constituents, particularly between head nouns and their modifiers.

[17] There is a range of academic genres in the academic corpus, including textbooks, journal papers, and monographs.

Although the quantitative comparisons were presented only by bar charts, the sharp contrasts of visualisation spelled out the register differences well.

Based on more rigorous statistical analysis, Biber et al. (2011) further called into question the appropriateness and adequacy of using subordination-based measures to study syntactic complexity in L2 writing. The rates of frequency (i.e. normalised frequency per 1,000 words) of 28 complexity features from three grammatical types (i.e. finite dependent clauses, non-finite dependent clauses, and dependent phrases) serving three grammatical functions (i.e. adverbial, complement, and noun modifier) in academic writing and conversation were compared for statistical significance (with ANOVA) and strength of association (using r^2). The results show conspicuous patterns for each grammatical type. For finite dependent clauses, most features are significantly more frequent in conversation than in academic writing, except for four features: predicative adjective + *that* complement clause (Ex. 23), noun + *that* complement clause (Ex. 24), and *which*-relative clauses (Ex. 25) which see significant greater frequency in academic writing, as well as *that*-relative clauses which see no significant difference.

(23) *Because the learners here appeared reluctant to make emotionally risky moves,* ***it is unsurprising that non-analytic talk occupied large amounts of small-group time***.

(24) *These differences would also explain why the results seemingly contradict* ***the view that the number of speakers affects the location of pairs across or within turns***…

(25) *Lexical cohesion concerns* ***the way in which lexical items (single words or multi-word units such as bed and breakfast and kick the bucket) relate to each other***.

For non-finite dependent clauses, five of seven features are significantly more frequent in academic writing: *to*-adverbial clauses (Ex. 26),[18] *adjective* + *-ing* clause (Ex. 27), adjective + *to* clause (Ex. 28), noun + *of* + *-ing* clause (Ex. 29), noun + *to* clause (Ex. 30), and non-finite relative clauses (Ex. 31).

(26) *This section first presents some numerical information and explanations for these results (see 'Frequency and Distribution of Lexical Cohesion' section), and then analyzes selected extracts of the conversations* ***to see how lexical patterns are collaboratively created and what goals are achieved***.

(27) *The content of their essays provided evidence that the students trusted each other and* ***felt comfortable sharing their life stories***.

(28) … ***it is perhaps safe to say*** *that what we are observing here is not an idiosyncratic phenomenon*…

(29) ***The importance of using questions*** *to check understanding is acknowledged in various teacher training texts in applied linguistics*.

(30) …*and* ***her attempt to check understanding*** *is received with laughter from the class (see transcription notations in the appendix*.

(31) *Classroom observation schemes often include* ***components targeting teacher questions with broad concerns***….

Finally, for dependent phrases all but one feature (i.e. adverbs as adverbials) are significantly more frequently used in academic writing. The features used more frequently in academic writing with an r^2 score higher than 0.50 (meaning that over 50% of the variation in the feature can be predicted by the register difference) are almost completely related to dependent phrases, representing the strongest structural type in

[18] Biber et al. (2011) did not include participle adverbial clauses to be analysed for non-finite adverbial clauses, but the authors did not explain why this was the case.

academic writing. The next most prominent structural category in terms of frequency is non-finite dependent clauses. The grammatical type that is least frequently used in academic writing is concerned with finite dependent clauses, thus characterised by the authors to least represent the syntactic complexity of academic writing.

These findings further reject the stereotype that the syntactic complexity of academic writing is elaborated in using dependent clauses but at the same time give a finer-grained, multidimensional picture of how different clause types contribute to the complexity of academic discourse. In terms of the distinction of finite and non-finite clauses, it has been found that different from finite clause types, non-finite clauses are favoured by academic writers. In terms of the three grammatical functions of clauses, they have also shown their own patterns of representation in academic texts. Based on the frequency results, finite adverbial clauses are more characteristic of conversation while non-finite adverbial clauses are preferred in academic writing.

However, participle adverbial clauses were not included for analysis of non-finite adverbial clauses. If they had been, they would probably also have shown greater frequency in academic writing as well. For complement clauses, it is interesting to note that complement clauses that are controlled by adjectives and nouns (Ex. 28-30), both finite and non-finite, are more characteristic of academic writing than those controlled by verbs, which are more frequent in conversation. Lastly, the use of relative clauses is more characteristic of academic writing than speech, although finite relative clauses introduced by the relativiser *that* have similar frequencies in the two registers and those by *which* are significantly more frequent in academic discourse.

The corpus evidence reported above has more empirically confirmed and further detailed functional linguists' theoretical characterisations of academic writing in contrast to speech. First, the linguistic features that most strongly characterise

academic writing all have to do with the expansion of the noun phrase: i.e. structures serving as pre- and postmodifiers of the head noun. This finding supports Chafe's and Halliday's discussion of nominalisation as being able to condense and pack information that is otherwise expressed through clauses in speech. Second, the linguistic features that least represent academic writing but best represent conversation are all clausal features, most of which have to do with finite adverbial clauses (hypotaxis) as well as complement clauses controlled by verbs (embedding). However, clausal features that are frequent in academic writing are mostly relative clauses and complement clauses controlled by nouns and adjectives (embedding). This finding is also consistent with Halliday's observation that speech is grammatically "intricate" through hypotactic clause chaining and the use of embedded clauses is more typical of writing.

It can thus be seen that the distinction between hypotaxis and embedding made by SFL scholars is a valid one since it can help identify register-specific category of clause combination, which the apparatus of *subordination* as defined in traditional grammar cannot accomplish.

2.4.2.6. Frequency Criterion for Determining Syntactic Complexity

This section discusses the extent to which it is appropriate and adequate to use frequency as the sole criterion to evaluate linguistic features related to syntactic complexity. The discussion focuses on some of the linguistic features that are significantly infrequent in academic writing and tries to provide some justifications for these features to be treated as playing important roles in academic writing as well. It first examines different discourse and meaning making roles of these structures in academic writing and then critiques the criterion of using frequency alone to determine syntactic complexity.

First, although found to be significantly less frequent in academic writing than in speech, finite adverbial clauses are used differently in the two registers, which has already been discussed in length in 2.4.2.4. To recapitulate, adverbial clauses in academic writing are needed to contribute to the argumentative network by encoding adequate and complex reason, justification, concession, and counterargument, in relation to the claims made in their matrix clauses. In contrast, conjunctions in conversation play a primarily pragmatic and interactional role as discourse markers guiding the listener to pay attention. Therefore, comparing the frequency of a clause type bearing radically different discourse functions and ideational complexity between two modes of communication may lose out on its important features in academic writing. As Schleppegrell (1996, p. 273) pointed out, "it is only by examining the ideational content of the clauses, the sequential distribution of the conjunctions, and the interactional contexts in which they occur that we can identify the functions they perform and the meanings they contribute".

In addition, although verb-controlled complement clauses have been found to be more frequent in speech than academic writing, one type of complement clause in fact plays a significant role in academic writing: complement clauses taken by reporting verbs. As acknowledged by Biber and Gray (2010, p. 9), the high frequency of verb-complement clauses in conversation is due to the frequent use of a set of lexically-restricted reporting verbs: *think*, *say*, and *know* account for around 70% of all occurrences of *that*-complement clauses in conversation; however, there are over 200 different verbs that can control this type of complement clause in all registers. Since an essential part of academic writing is to report previous research and results, it is necessary for researchers to use complement clauses as reported speech and thought. Moreover, academic writing uses a far wider range of reporting verbs than speech that

are varied in shades of meaning and evaluative stances (Thompson & Yiyun, 1991). The ideational meaning of the complement clause would be somewhat conditioned and guided by the specific reporting verb selected by the author.

Thus, similar to the differences of adverbial clauses in speech and writing discussed previously (Schleppegrell, 1996), the same can be said about the use of reporting verbs and complement clauses in the two modes of communication. Reporting verbs like *say* and *think* in conversation, with minimal semantic delicacy, play a primarily discourse-marking function in drawing the listener's attention to the speaker's narration or attitude. By contrast, reporting verbs used in academic writing introduce and interact with more complex information. To summarise, even though verb-controlled complement clauses are used significantly less frequently in writing than in speech, it would be problematic to leave out verb-controlled complement clauses as insignificant for academic writing.

While frequency data is able to delineate a broad syntactic profile of academic discourse, it would be questionable if frequency alone can characterise its syntactic complexity. High-frequency linguistic features are definitely prominent for noticing and should be maximally employed by the writer to achieve genre and style expectations, but they do not necessarily constitute *complexity*, which is understood to be the variety and sophistication of grammatical structures evident. Also, in terms of the relationship between frequency and complexity, it would be intuitive to posit that high-frequency items are usually from among the simpler, easy-to-use structural repertoire.

On another note, the frequency-based comparison of the totality of a particular linguistic feature without considering its different lexicogrammatical environments in speech and writing seems somewhat unfair. Since it has been agreed upon that grammar

is not easily separable from lexis, counting the frequency of a grammatical feature without considering its lexical accompaniment may distort claims made on it. For example, although finite adverbial clauses introduced by *because*, *if*, and *though* were found to be significantly infrequent in academic writing, other conjunctions more prominent in writing were not included, e.g. *since* and *as* for encoding reason, *while*, *whereas*, and *despite (the fact) that* for concession. Another example that has been mentioned before is that the set of verbs, adjectives, and nouns that can control finite complement clauses is much richer in writing than that in conversation. Therefore, if the frequency comparisons had been made in association with their lexical attributes, the results could have been different.

2.4.3. Summary of Syntactic Complexity of Academic Writing

The review has by far identified important grammatical features of academic discourse through examining theoretical discussions by functionally oriented linguists and empirical evidence offered by corpus studies. The most salient grammatical feature has been found to be concerned with condensing and integrating information, by which verb-centred processes are compressed into noun phrases which become nominalised participants of a new clause. Ideational complexity increases in such a process, in which nominalisation and grammatical metaphor play a crucial role. However, from a syntactic perspective, a nominalised grammar brings about a tendency towards simplicity rather than complexity in that the process of transforming finite clauses into noun phrases reduces the number of clauses overall. However, it has emerged that academic writing is also characterised by certain clausal structures as revealed by contrastive corpus analysis. Specifically, non-finite adverbial clauses and complement clauses controlled by adjectives and nouns are significantly more frequently used in

academic writing. It has also been argued that finite adverbial clauses, although found to be infrequent in academic writing, should be understood to have different discourse functions and encode different modes of meaning in academic writing. Likewise, verb-controlled finite complement clauses used with reporting verbs are arguably an essential clause type used in academic writing, particularly when the rich set of reporting verbs is considered. In addition, as noted by Chafe (1982), non-finite participle clauses (including verbless adjuncts) are another prominent clause type used in writing serving a number of grammatical functions, but unfortunately they were not included in Biber et al.'s (2011) analysis.

By now, a distinction between noun phrase complexity and clausal complexity as playing different roles in academic writing can be made. Noun phrase complexity is attributable to the use of noun phrases with complex modifications to efficiently integrate information, whereas clausal complexity emerges as clauses composed of complex noun phrases interact with each other to achieve argumentative aims. Although it is certain that noun phrase complexity leads to the reduction of the use of verbs and hence clauses in the overall discourse, the importance of clause types with writing-specific lexicogrammatical features should not be ignored. The two dimensions of syntactic complexity do not have to conflict with each other, but rather exist to achieve different purposes: noun phrase complexity for integrating and sequencing processes and entities, clausal complexity for imposing argumentative significance on them. With syntactic complexity also referring to the variety of syntactic structures used in text generation, the rich array of clause types available in English grammar should find their way into the examination of syntactic complexity of academic writing.

The next section will review research on the syntactic complexity of second language writing, aiming to find out aspects of syntactic complexity as examined above that are emphasised by L2 writing researchers.

2.5. Syntactic Complexity in L2 Writing Research

In SLA, linguistic complexity is one of three frequently studied constructs of L2 proficiency and L2 performace, together with accuracy and fluency, often shortened as the acronym *CAF* (Housen & Kuiken, 2009; Skehan, 1989). Complexity appeared in the 1990s as an addition to the classic dichotomy of accuracy and fluency (Brumfit, 1984). Empirical findings suggest that these three constructs are often interdependent and compete with each other. For instance, speakers' fluency may be affected by conscious attention to forms and the process of monitoring output which contribute to accuracy and complexity. In writing, where fluency is a less prominent concern, creating grammatical complexity may be correlated with reduced accuracy. However, the biggest challenge for advanced L2 writers, who are arguably less likely to be greatly troubled by grammatical errors than less proficient writers, would be the need to codify complex ideas and arguments in rhetorically and pragmatically appropriate ways.

Syntactic complexity, together with lexical complexity, has been regarded as an important indicator to gauge overall L2 writing development and has generated much work in L2 writing research. As such, the original purpose of conducting syntactic complexity research is not to examine learners' writing proficiency, but to monitor development. However, the underlying assumption about syntactic complexity that mature writing should show a variety of syntactic structures can also lend itself to the understanding of written product as evidence of proficiency. Although much research on syntactic complexity tracks students' writing development across curriculum or

proficiency levels (Lu, 2011), there are also studies that explore the phenomenon cross-sectionally, e.g. between texts of different genres (Staples et al., 2016) and by writers of different L1s (Lu & Ai, 2015).

This section reviews commonly used measures of syntactic complexity in relation to their underlying grammatical significance, identifies measures that effectively distinguish curriculum levels, and discusses problems and challenges facing L2 syntactic complexity research. The review will be closely tied to previous discussion of the lexicogrammatical features of academic discourse.

2.5.1. Syntactic Complexity Measures

Central to the study of syntactic complexity in L2 writing is the use of measures (or metrics) for its quantitative analysis. Polio (2001) identified three main purposes of measures of syntactic complexity for use in L2 writing research: 1) to evaluate the effects of pedagogical treatment on the development of grammar and writing ability, 2) to investigate task-related variation in L2 writing, and 3) to assess differences in L2 texts by learners across proficiency levels and over time. A fourth purpose that has not been identified by Polio is to use syntactic complexity measures to examine differences in L2 texts written by learners with different L1 backgrounds to reveal cross-linguistic influences on syntactic complexity (e.g. Lu & Ai, 2015).

Slightly different from L2 writing research, the ultimate purpose of using complexity measures in the present study is to identify areas where various aspects of syntactic complexity are underdeveloped in student academic writing and instruction could be expected to be fruitful. Although L2 writing researchers have cautioned that syntactic complexity measures should not be used to measure or be equated with writing ability (Ortega, 2003; Wolfe-Quintero, Inagaki, & Kim, 1998), scholars more

concerned with EAP instruction take an approach to syntactic complexity that focuses on specific high-stakes linguistic features students need to develop for more effective writing (e.g. Biber et al., 2011; Parkinson & Musgrave, 2014). This view of syntactic complexity also provides impetus for the present study.

There are two general types of syntactic complexity measures: ratio measures and frequency measures. Ratio measures calculate the average number of a particular grammatical feature as a proportion to a production unit, or that of a lower-level production unit in relation to a higher-level unit. The production units used for the measures include clauses, T-units, and sentences. T-unit, or "terminable unit", is a term proposed by Hunt (1965) referring to an independent clause (i.e. main clause) together with any number of subordinate clauses appended to it.[19] It is equivalent to a sentence if the sentence consists of only one main clause, but it is smaller than a sentence if there is more than one main clause in it. Being potentially multi-clausal, the T-unit was designed to examine the elaborateness of subordination. Examples of ratio measures are T-unit complexity ratio (clauses per T-unit) and dependent clauses per T-unit. On the other hand, frequency measures calculate the average occurrences of a particular grammatical structure per number of running words. This is similar to the method used in corpus-based studies of grammatical features in academic writing reviewed above. In L2 writing research, ratio measures are more commonly used than their frequency counterparts.

Apparently, measurement of noun phrase complexity, which has been found to be the most distinctive feature of academic discourse, is absent from these two measure types. However, it should be noted that length-based measures include those measuring the length of a potentially multi-clausal T-unit or sentence as well as that of a clause,

[19] Subordinate clauses in writing research include both hypotactic and embedded clauses.

which may rely on different expansive mechanisms. The length of a multi-clausal T-unit or sentence may be realised through subordination and coordination, whereas the clause becomes longer only through the expansion of its nominal components. In this sense, the measures of the length of a potentially multi-clausal production unit gauge overall syntactic complexity whereas the measure of clause length can be regarded as tapping into noun phrase complexity.

Surveying the most frequently used syntactic complexity measures, Lu (2010) identified 14 measures subsumed under five types: 1) length of production unit, 2) sentence complexity, 3) subordination, 4) coordination, and 5) particular structures. Similar to Norris and Ortega's (2009) finding, length-based and subordination-based measures predominate, accounting for half of the 14 measures. However, Type 5, "particular structures", includes two measures concerned with noun phrase complexity: complex nominals per clause and complex nominals per T-unit. The feature complex nominal was used by Cooper (1976) to include three specific nominal structures: 1) nouns with possessive, adjectival and preposition phrase, participle, relative clause, or appositive, 2) nominal (or complement) clauses, and 3) infinitives and gerunds in subject position. It can be seen that only the first category is directly concerned with the use of noun phrases in its strict sense, with the other two structural types only sharing a nominal property. As such, the complex nominals measures can only partially capture the use of noun phrase complexity, calling for more straightforward metrics of how writers use complex noun phrases.

Norris and Ortega (2009) highlighted two problems concerning the conceptualisation and operation of the different measures: inadequate understanding of syntactic complexity and redundant measures. Not solidly based on linguistic insights into the lexicogrammatical features of academic writing as reviewed in 2.3, L2 writing

researchers have taken a more or less intuitive approach to syntactic complexity. That is, complex writing is characterised by long sentences and frequent use of subordinate clauses. As already made clear up to now, L2 writing researchers' sole belief in subordination falls foul of linguists' observation that it is much more of a feature in spontaneous speech than in academic writing, particularly in frequency terms. Biber et al. (2011) in particular take up this point to criticise L2 researchers' overreliance on subordination-based measures to study the syntactic complexity of EFL academic writing.

With respect to viewing complexity as embodied by length of production units, measures of length of production units risk being understood as dealing with the same underlying source of complexity. As already mentioned, a potentially multi-clausal production unit such as a T-unit or a sentence may become longer through the addition of subordinate clauses, but a single clause (regardless of their status as an independent or dependent clause) can only become longer through noun phrase complexity. Therefore, the results of length-based measures would represent different sources of syntactic complexity and may vary with such factors as proficiency, genre, or curriculum level.

To arrive at a more clear-cut operationalisation of syntactic complexity measures, Norris and Ortega (2009) categorised length-based and subordination-based measures into three sub-constructs: 1) overall or general complexity, measurable by length of a potentially multi-clausal production unit such as the T-unit, 2) complexity via subordination, measurable by any metric with clause (of any type) in the numerator, and 3) subclausal (or clause-internal) complexity via phrasal elaboration, measurable by length of clause and partially captured by metrics of complex nominals ratio. It needs to be reminded that the use of non-finite and verbless clauses (also called fragments

and adjuncts), usually coded as phrases in writing research, cannot be measured by subordination-based measures but can only be tapped via measures of phrasal complexity. Since these structural types are identical with neither finite clauses nor phrases, it would be necessary for them to be measured in their own right.

As already reviewed, both Norris and Ortega (2009) and Lu (2011) have found that the most frequently used measures centred around subordination-based complexity. By contrast, examination of phrasal complexity demonstrates a major underrepresentation. This unbalanced emphasis placed on different measure types reflects a biased understanding of the nature of syntactic complexity. Obviously, researchers have preferred to think of complexity in terms of amount of subordination involved in a potentially multi-clausal production unit, overlooking the multidimensionality of the construct. Norris and Ortega (2009), partly inspired by the notion of grammatical metaphor, proposed that phrasal complexity at the sub-clausal level should merit much more attention. Following their call, a growing body of research has emerged in the past few years that aims to chart cross-sectional variability of discrete complexity dimensions and longitudinal development trend of each dimension and the interaction among the dimensions (e.g. Biber et al., 2011; Byrnes, 2009; Byrnes, Maxim, & Norris, 2010; Parkinson, 2015; Parkinson & Musgrave, 2014; Spoelman & Verspoor, 2010; Vyatkina, 2012, 2013).

As for redundancy of measurement, Norris and Ortega pointed out that some metrics measure exactly the same thing and so are redundantly used. Among the many metrics developed for measuring subornation, most are premised on the same source of complexification, i.e. the addition of subordinate and embedded clauses onto a matrix clause. Examples of measures of subordination are T-unit complexity ratio (clauses per T-unit), complex T-unit ratio (complex T-units per T-unit), and dependent clause ratio

(dependent clauses per clause and dependent clauses per T-unit).[20] The authors concluded that using any one metric from the subordination family "is sufficient to gauge complexification that is achieved by means of subordination, and using more than one together for the analysis of the same data would be redundant" (ibid, pp. 560-561). The problem with redundancy, as the authors pointed out, is usually a consequence of researchers' lack of awareness of the multi-dimensional nature of syntactic complexity. However, it is understandable that more than one measure for the same complexity construct has been developed because researchers have been trying to find measures that more effectively chart developmental trajectories.

Although measures of syntactic complexity were initially developed out of largely intuitive conceptualisation, i.e. longer production and use of subordination, researchers have increasingly come to realise that syntactic complexity is derived from different sources of complexification and should be measured multi-dimensionally. The next two sections review how different aspects of syntactic complexity relate to important variables of student writing that may have impact on them. The specific variables to be examined include proficiency, writing quality, genre, and L1 background.

2.5.2. Developmental Trajectories of Syntactic Complexity in L2 Writing

Overall, the three main dimensions of syntactic complexity increase along with proficiency development, although subordination-based complexity has shown inconsistent results. Wolfe-Quintero et al. (1998) examined the strength of the relation between syntactic complexity measures and proficiency levels in 39 college-level L2 writing studies. They found that mean length of T-unit, mean length of clause, clauses

[20] See Lu (2010) for source information of the measures.

per T-unit, and dependent clauses per clause were the best measures that could distinguish proficiency, curriculum, and holistic rating levels. In other words, the sub-constructs of syntactic complexity underpinning these measures increase along with writers' language proficiency and general writing quality development. Thus, overall complexity (as measured by mean length of T-unit), subordination-based complexity (as measured by dependent clauses per T-unit and dependent clauses per clause), and phrasal complexity (as measured by mean length of clause)—they all showed linear growth. This finding is consistent with Ortega's (2003) research synthesis of 25 college-level L2 writing studies, except that subordination has shown both positive and negative changes across the studies, indicating that for certain groups of L2 writers the use of subordination decreases at certain points of proficiency development.

Byrnes et al. (2010) provide a detailed deconstruction of discrete developmental pathways of the three complexity dimensions of overall complexity, subordination complexity, and phrasal complexity (respectively embodied in the measures of mean length of T-unit, clauses per T-unit, and mean length of clause) for advanced learners of German whose L1 is English. Both cross-sectional and longitudinal findings present a similar developmental trend for the three dimensions. Cross-sectional results show that across the four curricular levels 1) there was a positive significant increase in overall complexity at each adjacent interval; 2) subordination-based complexity showed continuous increase at the first two intervals but a drop at the last one; and 3) increase in phrasal complexity also occurred at each adjacent interval except between Level 1 and 2 (almost no difference at this interval).

The same pattern holds for longitudinal results only except that there is no increase in subordination at the second interval, i.e. between Level 2 and 3. Thus, it could be said that increase in overall complexity is attributable to growth in both phrasal

and subordination complexity at the lower curricular levels, i.e. Level 1 and 2, possibly more by subordination. In contrast, at the more advanced levels, overall complexity increase could mainly be a result of greater phrasal complexity, and, correspondingly, a shrinking subordination employment. By the same token, Lu (2011) found that out of 14 frequently used syntactic complexity measures the best candidates that discriminated proficiency levels were complex nominals per clause (CN/C) and mean length of clause (MLC), both of which tap clause-internal phrasal complexity.[21]

In contrast to Byrnes et al.'s (2010) investigation of advanced German L2 learners, Vyatkina (2012, 2013) provides complementary results obtained from beginning level learners concerning the longitudinal development of specific complexity dimensions. Similar to the findings for advanced groups, Vyatkina (2012) also found an increase in overall complexity as measured by sentence length (from 6 words per sentence to 10 words). But different from Byrnes et al.'s finding that the overall complexity of advanced German L2 learners resulted more from greater phrasal complexity rather than subordination, gains in overall complexity by Vyatkina's beginner-level learners were mainly due to increased subordination use (see Vyatkina, 2012, pp. 584–588 for detailed results). This contrast may help to map out the developmental sequence of specific complexity dimensions along the proficiency or curriculum scale. For clause-internal phrasal complexity, it seems that a major growth would not occur until more advanced stages of literacy development.

Thus far, it can be said that the developmental trajectories of subordination complexity and phrasal complexity have shown a dynamic relationship between the two dimensions. Ortega (2003) and Byrnes et al. (2010) have identified a "trade-off"

[21] A good measure is one that can yield linear progressive complexity values significantly correlated with curricular levels.

effect in this relationship, particularly towards higher proficiency and literacy levels. That is to say, as learners become more proficient overall, they tend to use more phrasal structures in lieu of clausal structures. As noted by Ortega (2003, p. 514), this evidence empirically supports the developmental prediction proposed by Cooper (1976) and more recently Wolfe-Quintero et al. (1998). The prediction argues for a non-linear development of subordination, contending that advanced proficiency groups would depend on phrasal, rather than clausal, resources in written production. This non-linearity has also been attested by research on L2 proficiency development approached from the perspective of language as a complex dynamic system (Larsen-Freeman, 2006, 2009, 2011).

This developmental proposal finds theoretical affinity with SFL's characterisation of the lexicogrammatical differences between speech and writing, where speech is characterised by clause chaining and writing by information packing by phrasal structures. Within this affinity, written product by lower-proficiency groups would bear much resemblance to speech, while that by higher-proficiency writers take on more features of mature academic writing. Therefore, the linear progression towards phrasal complexity and a later decrease in subordination use indicate that second language writing gradually moves away from an oral style en route to the target academic register. Indeed, the oral tendency in L2 writing has been captured by prior research with respect to the use of a wide range of speech-like lexical and grammatical features (e.g. Hinkel, 2002; Gilquin & Paquot, 2008).

However, despite the finding that subordination complexity shows a decrease towards higher proficiency levels, from a usage perspective it is not really known which aspects of subordination have dwindled and if the decrease takes place across all subordinate structures. Since the construct of subordination in L2 writing subsumes

both hypotactic and embedded clause types, it would be unlikely to distinguish their own developmental trajectories and pinpoint the difference, if any. Predictably, (restrictive) relative clauses which directly contribute to noun phrase complexity and complement clauses which have a nominal property would see an increase, rather than a decrease, as proficiency grows. This trend is not inconsistent with the developmental prediction because of these clause types' implications for noun phrase complexity. Even for adverbial clauses, which have been labelled as a feature of orality in frequency terms, the discourse function and semantic force of adverbial clauses characteristic of academic discourse emerge as proficiency grows. In other words, adverbial clauses with speech-oriented functional and semantic features should be expected to decrease while those featuring writing-specific features would take on greater presence (cf. Schleppegrell, 1996).

Although longitudinal studies have identified a decrease in subordination and an increase in phrasal structures, cross-sectional studies comparing the writing by native and non-native speakers or by low- and high-proficiency writers have consistently found that the more advanced writers use both more subordination and phrasal structures than the less developed writers (Ai & Lu, 2013; Lu & Ai, 2015). These findings cast doubt on whether subordination continues to drop as proficiency and literacy grow to even more advanced levels and call for continued research in this area. This has important implications for the present research which is set to examine whether postgraduate students still need to further develop subordination use in academic writing.

To compound the situation, not much is known about the developmental trend of non-finite clauses of all types in that they are not included in the coding of subordination in L2 writing research. With non-finite clauses coded as non-finite verb

phrases, their representation is hidden in the length of (finite) clause. Thus, the measure of clause length would somewhat obscure the significance of different structural types that can be included in what is defined as a finite clause. Neither the subordination-based measures nor the measure of clause length is able to clearly address the complexity associated with non-finite structures, calling for more specific measures tapping this unique structural type. These differentiations are necessary if syntactic complexity research is to be based on functionally meaningful grounds (Ortega, 2015).

2.5.3. Syntactic Complexity and Other Variables

2.5.3.1. Impact of Genre on Syntactic Complexity

This section briefly reviews the relationship between syntactic complexity and quality of writing. However, the identification of such a relationship needs to consider the specific genre under examination because different genres would have more or less different expectations for lexicogrammatical configuration. Homogeneous genres traditionally includes such modes of discourse as narration, description, exposition, and argumentation. Since most studies on L2 syntactic complexity have been based on writing samples taken from one of the above genres, the findings and claims derived from these studies can only be best interpreted from the perspective of the particular genre in question. For example, investigation into the developmental trajectories of syntactic complexity reported above has seldom differentiated variability caused by differences in genre. In fact, the impact of genre has only begun to be realised by researchers interested in syntactic complexity, who bring genre differentiation as a critical variable moderating syntactic complexity (e.g. Biber, Gray, & Staples, 2014; Staples, Egbert, Biber, & Gray, 2016; Beers & Nagy, 2009, 2010; Berman & Nir-Sagiv, 2007; Crowhurst, 1980a). As can be seen from the citations, much of the research in

this area takes place in the study of syntactic complexity of first-language writing, calling for more focused effort in this regard from L2 writing researchers.

While there could be distinct lexicogrammatical landscapes of homogeneous genres, it would be difficult to assess how the impact of genre works for examining the syntactic features of such heterogeneous genres as research articles and dissertations. This is so because the research article or dissertation is constructed by multiple modes of discourse characterising different rhetorical sections (e.g. introduction, methodology, and conclusion). However, exposition and argumentation, rather than narration, are the dominant modes. In this sense, findings derived from syntactic complexity research based on argumentative and expository texts can still be expected to shed some light on the research programme of this study.

2.5.3.2. Syntactic Complexity and Writing Quality

Studies examining the relationship between different aspects of syntactic complexity and overall writing quality have shown a largely positive correlation, although with variability in the interaction between measures and genres (Beers & Nagy, 2009; Bulté & Housen, 2014; Crowhurst, 1980b; Yau & Belanger, 1984). In this type of correlational research, writing quality is usually determined by teachers' or researchers' impressionistic ratings of written product based on holistic scoring rubrics, without raters' attention explicitly directed to the manifestation of syntactic complexity. Bulté and Housen (2014) found that growth in T-unit and clause length showed good correlations (r = .40 and .48 respectively) with writing quality ratings. However, this correlation varies when examined across different genres. For example, Beers and Nagy (2009) found that in L1 English high-school students' writing, mean length of clause (a measure of phrasal complexity) was positively correlated with writing quality for

argumentative essays only, while mean clauses per T-unit (a measure of subordination) was positively correlated with quality for narrative essays and negatively correlated with that of argumentative writing. Similarly, an early study by Crowhurst (1980b) showed evidence that L1 English pupils' argumentative texts of high syntactic complexity determined by T-unit length (a measure of global complexity according to Ortega, 2003) were rated significantly higher than those of low complexity, whereas narrative texts of high syntactic complexity were rated significantly lower than those of low complexity; in this study, statistical significance occurred only for writing at higher grade levels (Grade 10 and 12). These findings seem to suggest that high-quality argumentation is more dependent on phrasal complexity but less on subordination, with the latter being a feature of high-quality narration. Since academic writing is mainly composed of the expository and argumentative genres, it would be reasonable to expect greater phrasal complexity used in high-quality academic writing as well.

While phrasal complexity from the use of complex noun phrases has been found to correlate with high-quality academic writing, it has also been found that its representation in the syntactic slot of grammatical subject (as opposed to logical subject) is particularly prominent and able to differentiate different writing quality (e.g. Callies, 2013; Lu, 2011; McNamara, Crossley, & McCarthy, 2009). McNamara et al.'s (2009) syntactic complexity measure of length before the main verb was one of three most predictive indices of quality of essays written by American university students. Theoretically, elements in a finite clause before the main verb can be grammatical subjects (mandatory) and adverbials (optional) which can be realised by preposition phrases or disjuncts (i.e. sentence adverbs such as *however*). Therefore, lengthy elements before the main verb are most probably subjects and preposition-phrase adverbials, the former of which can be realised by what is termed "complex nominals"

in writing research, i.e. complex noun phrases, nominal clauses, or gerunds and infinitives (Cooper, 1976). Frequency-wise, however, it is the use of noun phrases that is the staple structure to realise subjects, playing the most significant role in McNamara et al.'s finding.

Nominal clauses and gerunds/infinitives serving as subjects are also significant mechanisms to add to text quality, as supported by the findings in Lu (2011) who found the two measures of complex nominals to best distinguish writing at different curriculum levels. Similarly, Callies (2013), comparing learner writing with that of native speakers, found that learners used significantly fewer inanimate/abstract subjects than native speakers, instead overusing first person pronouns and subject placeholders (e.g. dummy *it* and existential *there*). The strong correlation between the use of lengthy NP-subjects and writing quality is partially attributable to its ability to create cohesion and coherence as important components of high quality writing. Vande Kopple's (1994) textual and sociolinguistic explanations of academic writers' preference for extended NP-subjects discussed above may lend solid support for this correlation.

2.5.3.3. L2 Syntactic Complexity and Writer's L1

While the influence of the lexical and grammatical features of the L1 on second language learning in general has long been studied and documented, L1 impact on the syntactic complexity features of second language writing has only begun to be recognised (Lu & Ai, 2015; Ortega, 2003). Ortega (2003) draws attention to the effect of cross-rhetorical transfer in assessing the syntactic features of students from different L1 backgrounds. For instance, Neff, Dafouz, Díez, Prieto, and Chaudron (1998) found that Spanish EFL writers showed a pronounced preference for subordination than phrasal elaboration than L1 English writers and that their level of subordination use

was similar to L1 Spanish journalism writing. Cross-rhetorically, this preference could have been transferred from the rhetorical convention of using subordination in Spanish, given the commensurate level in L1 Spanish discourse. In a major study examining the L1 effect on syntactic complexity in college-level writing among writers with different L1 backgrounds, Lu and Ai (2015) found significant differences between the NS group and one or more NNS groups in all 14 measures of syntactic complexity tested, as well as among NNS groups from the same proficiency level. Their findings revealed that Japanese and German writers respectively represented the lowest and highest level of syntactic complexity for most of the measures used. Notably, L1 English writers (i.e. US college students) did not score the highest for any of the 14 measures, roughly ranking at an upper-middle stratum. As the authors argue, "intergroup variation in syntactic complexity cannot be accounted for by proficiency alone but that learners' L1 may play a role in the syntactic complexity in their L2 writing as well" (p. 24). Given the findings of their study, this role should be a massive one in shaping the syntactic contour for writers from a particular L1 background.

2.6. Syntactic Complexity of Chinese EFL Writing

While there is a large body of research into Chinese EFL writing, not much focus has been placed on investigating its syntactic features at a particular curricular level or development over time. Information on the syntactic features of Chinese student writers has been mostly gleaned from studies in which Chinese students constitute a part of the L2 population under examination. This section first reviews general syntactic features of Chinese students' EFL writing before going on to examine its syntactic complexity.

2.6.1. General Syntactic Features of Chinese Students' Writing

Hinkel (2003) labelled ESL academic texts written by Chinese (among other Asian students) as showing "simplicity without elegance" (p. 275), possibly alluding to the Hemingway style of economical and laconic journalism. Instead of focussing on the extent to which student writing exhibits features of syntactic complexity, Hinkel examined the frequency rates of what he termed simple syntactic and lexical features in ESL writing. The syntactic features Hinkel argues to make academic writing appear simple are all centred around the use of copular BE and its attendant structures including predicative adjectives and existential *there*-constructions. By contrast, the *it*-cleft construction used together with BE is seen as "advanced and sophisticated" (p. 281), compatible with the formal and depersonalised academic style. Hinkel's corpus-based analysis showed that Chinese students used significantly more copular BE and predicative adjectives but significantly fewer *it*-cleft constructions than native speakers. However, the Chinese students' texts did not score the lowest on these features, with all the other L2 groups (including Japanese, Korean, Indonesian, and Arabic speakers) demonstrating the same pattern by comparison with the native-speaker group.

Despite Hinkel's finding that overuse of BE is responsible for an unsophisticated and spoken syntactic contour in Chinese students' writing, this in fact represents more of a lexical than a syntactic attribute, apart from being not capable enough to address the syntactic features that can grapple with the awkward simplicity of L2 writing. On the one hand, the structure copular BE + predicative adjective, although basic and simple, is an important discourse unfolding strategy in academic discourse. Occurring approximately 8,500 times per million words, it is more frequently used in academic writing than in other registers (Biber et al., 1999, p. 437). The problem with students' use of this structure, as can be seen from the essay extracts provided in

Hinkel (2003), is not so much a matter of using too much of this structural type as one of overusing animate subjects and adjectives expressing human emotion and condition (e.g. *sleepy*, *enjoyable*, and *friendly*). In other words, the discourse semantics mapped onto this structure is an informal, spoken one, the use of the structure itself not to blame.

On the other hand, the extraposed *it*-cleft structure does not seem to present great difficulty for acquisition given its clear-cut syntactico-semantic correspondence. Once learnt, it may run the risk of being overused. As Callies (2013) found, academic writing by advanced EFL students used significantly more *it*-cleft constructions as well as existential *there*-structures than L1 English writers, which accounted for significantly underrepresentation of inanimate and abstract subjects. Again, using cleft structures does not necessarily enhance syntactic sophistication, which is much dependent on the content of message conveyed within clauses. This brings the discussion back to the importance of clause-internal information condensation, i.e. noun phrase complexity, as the most fundamental feature giving rise to syntactic complexity.

2.6.2. Syntactic Complexity of Chinese EFL Writing

Most research on the syntactic complexity of Chinese students' EFL writing is based on argumentative essays written at the undergraduate level, and mostly conducted with a corpus-based approach (Ai & Lu, 2013; Lu, 2011; Lu & Ai, 2015). As discussed in 2.3, little effort has been made to address the syntactic complexity features of postgraduate academic writing by either Chinese or other L2 students. However, research on undergraduate essays can also provide useful information on the syntactic complexity features of Chinese students' writing.

Ai and Lu (2013) and Lu and Ai (2015) examined the syntactic complexity of Chinese writers by comparison with native speakers and other L2 students, whereas Lu (2011) investigated developmental trends of different dimensions of syntactic complexity to identify the best measures to differentiate among L2 texts at distinct curriculum levels. Ai and Lu (2013) compared argumentative and expository writing by Chinese college students with essays (genre not clear) written by L1 English students and found significant differences between the two groups of students in all four dimensions of syntactic complexity; however, they found no significant differences in two measures of clausal coordination between Chinese students and native speakers, which the authors ascribed to early instruction of this linguistic feature. Lu and Ai (2015) found that among argumentative essays written by students from 8 L1 backgrounds (including English) Chinese college students' syntactic complexity ranked the lowest for 5 measures and 2nd lowest for 3 measures, only slightly higher than that of Japanese students' essays. It needs to be noted that the lowest-scoring measures for the Chinese group are all concerned with subordination and coordination, devices that are considered to characterise speech rather than writing. Interestingly enough, the two measures on which Chinese students scored the highest were measures of noun phrase complexity.

Since the proficiency level of the Chinese students was rated as the lowest among the eight groups, it would be difficult for their strong performance in noun phrase complexity and weakness in clausal complexity to be explained in light of the developmental prediction. This is because this prediction states exactly the opposite, i.e. use of clausal structures is associated with early development while that of phrasal structures with more advanced literacy. Therefore, the fact that the Chinese students' less advanced proficiency has demonstrated a relatively high level of phrasal

complexity should be interpreted from other perspectives, for example the cross-rhetoric prediction. Although NP modification in Chinese is only prenominal, it is quite common to use multiple premodification in Chinese academic writing as in English and both phrasal and clausal premodifiers are possible. Thus, Chinese students would not feel uncomfortable to construct complex noun phrases in English academic writing.

By the same token, the cross-rhetoric predication can also to some extent explain the extremely low representation of Chinese students' use of subordination and coordination. While the rhetorical tradition of the ancient eight-legged essay no longer has a great impact on modern Chinese prose as it is primarily based on the western format (Mohan & Lo, 1985), the clause- and sentence-level syntax and inter-sentential semantics of the Chinese language (e.g. Chan, 2004) and discourse (e.g. Plaks, 1988) do greatly influence Chinese students' English texts. Among other things, the major source of influence comes from the preference of Chinese discourse for parallelism and loose semantic relations among statements (Kaplan, 1966). One of the syntactic consequences of this rhetorical feature is implicit, hidden coordination and subordination among phrases and clauses (e.g. Cheng & Wang, 2004; Lu & Ai, 2015). Thus the relationships between statements, if any, are usually left to the reader, as Chinese rhetoric has often shown to be reader- rather than author-responsible. These rhetorically-based underpinnings of Chinese discourse construction could possibly explain the low representation of subordination in Chinese students' writing.

However, the developmental patterns of Chinese student writing seem to support the developmental prediction (Lu, 2011). Lu found that measures of subordination showed non-linear, curvilinear development across proficiency levels, with an initial increase and a linear decline along higher levels, whereas measures of clausal complexity increased linearly across all proficiency levels. As such, despite

Chinese students' relative familiarity with noun phrase structures and culturally and rhetorically driven refrain from subordination by comparison with other L2 writers, their developmental trajectory still follows a pattern common to all L2 writers. But the vexed question is whether Chinese students' already underrepresented subordination may continue to decline as they capitalise more on phrasal structures, probably to a point where adequate inter-clausal meaning making cannot be achieved. The present research, which examines Chinese postgraduate students' academic writing in comparison with expert writing, can provide further information on whether highly advanced Chinese (and of course other L2) writers need to enhance or rein in subordination use.

2.7. Summary of Chapter

This chapter has provided a comprehensive review of the major issues related to the examination of syntactic complexity of EFL dissertation writing. The review first examined existing research into dissertation writing and found that most attention has been put on studying its generic features such as organisational structure. It went on to synthesise results from surveys of dissertation writers concerning their perceived challenges in the writing process and concluded that students generally see linguistic problems as more difficult to handle than organisational concerns. This finding further legitimates the need for close attention to linguistic barriers facing research students, although they may be considered to have already reached an advanced literacy level. Next, the linguistic features of academic discourse were examined from the perspectives taken from functionally oriented linguistic theories and came up with the consensus that a nominalised grammar dominates academic writing with a heavy reliance on complex noun phrases used in metaphoric, figurative manners as the major

source of text complexity, for both production and comprehension. This claim is empirically supported by findings from corpus linguistic investigations of academic discourse in comparisons with conversational features.

However, a case has been made for the importance of subordination—or rather, clause combination—in academic writing by demonstrating its different functional roles and semantic attributes in speech and writing: while clause combination in speech is universal and conjunctions mainly serve interactional, pragmatic functions, that in academic writing—although much less frequent—is used to advance complex argumentation and conjunctions are used to mark different stages of this advancement.

Finally, the issue of syntactic complexity was reviewed as a major topic in second language writing research, with researchers gradually coming to realise the relative importance and developmental patterns of phrasal and clausal complexity and that syntactic complexity is multidimensional and determined by such factors as genre and L1. Furthermore, it also behoves researchers to expand the horizons of L2 syntactic complexity research by exploring the usage-based and functional significance of changes in syntactic complexity measures. A close examination of what motivates and constrains L2 students to produce ideationally complex language is expected to take centre stage in this usage-based turn.

Thus far, the literature review has provided a useful research basis for the present study. At the same time, important research gaps and underexplored areas have been revealed, which stand out to motivate the present study. On the one hand, syntactic complexity, although having a long research tradition in second language writing, has been found severely underrepresented in research on L2 dissertation writing. Since the 1990s, the focus of L2 writing research has shifted from undergraduate essay writing to postgraduate disciplinary writing as a result of increased interest in disciplinary

genres (Leki, Cumming, & Silva, 2010, p. 42). Against this paradigm shift, it is necessary for important research topics in L2 writing such as syntactic complexity to be incorporated into L2 dissertation writing. Practically speaking, developing syntactic complexity is also crucial for EFL research students, for whom one of the greatest concerns in their research studies is the disheartening mismatch between disciplinary expertise growth and inadequate familiarity with complex meaning-making devices valued in academic discourse; this practical side no doubt necessitates a study like the present one.

On the other hand, the need for finer-grained analyses of different dimensions of syntactic complexity over and above mere quantification by syntactic complexity measures is expanding. Both features of phrasal compression—noun phrase complexity in particular—and those of clause combination merit much more detailed cataloguing and analysis than is presently dominant in L2 syntactic complexity research. This line of enquiry also calls for functionally significant linguistic theories such as SFL and other culturally and contextually informative models such as contrastive rhetoric (Connor, 1996) and argument structure (Crammond, 1998) to join forces to provide meaningful support and explanations for surface structural manifestations (see Section 2.5 and 2.6).

After taking stock of the major issues relevant for this study, the next chapter will present the methodological framework and detailed analytical procedure for the different aspects of syntactic complexity reviewed in this chapter. To answer the three research questions raised in 1.6, the research design will map out three independent, but related studies to address the three questions, which are arranged in three chapters.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

3.1. Overview of Methods and Methodologies in L2 Writing Research

Hyland (2016) provides a comprehensive introduction to the major methods and methodologies in second language writing research. Methods, as Hyland points out, are the ways of collecting data and include elicitation, introspection, observation, and text sampling. By contrast, methodologies are "the principles and understandings that guide and influence our choice and use of methods" (ibid, p. 177), including ethnography, auto-ethnography, critical analysis, discourse analysis, and meta-analysis. The kind of data that interests researchers is determined by their preconceptions and assumptions of what writing is and how it might be best understood. In the same vein, the selection of methodologies is closely associated with our philosophical beliefs about the world and possible ways of knowing it.

Hyland (2016) summarises six broad views of writing, each of which can be associated with one or more data collection methods and research methodologies: writing as 1) expressive activity, 2) cognitive activity, 3) completed activity, 4) situated activity, 5) social activity, and 6) ideology. To a greater or lesser extent, these views of writing can be distinguished by the larger dichotomy of writing as process and writing as product, as introduced in Chapter 2. Except for the view of writing as completed activity which emphasises formal features of texts isolated from socio-cultural and cognitive underpinnings, all the other perspectives take writing as part of socially or cognitively constructed processes. Traditionally, the study of syntactic complexity in L2 writing has employed data collection methods and analytical procedures that take

student writing as a largely isolated object to be measured, i.e. completed written product as a result of teaching and learning. Thus, notions like genre, discipline, topic familiarity, cognitive maturity, writer identity, and communicative purpose are often not considered when examining syntactic complexity. However, this research domain in L2 writing is currently undergoing a cognitive and functional turn (Ortega, 2015), along with growing awareness of the need to explore the factors that impact on formal features.

While the present study also examines student writing as informed by the writing-as-product paradigm, the motivation behind it comes from a primarily usage-based belief. Therefore, the measurement and analysis of students' written output will be grounded in its functional, cross-rhetorical, cognitive, and disciplinary underpinnings. In other words, the various processes joining force to form the linguistic features of student writing will be constantly examined to better understand the driving forces of syntactic complexity (or the lack thereof) in student writing. Notably, as will be seen, this study controls the genre (i.e. empirical academic writing) and discipline (i.e. applied linguistics) of student writing in the first place, allowing for more meaningful and contextualised results of comparison and their interpretation.

3.2. Contrastive Approach to Text Data

This study takes a contrastive (or comparative) approach to student academic writing, as does most L2 writing research employing the textual analysis methodology. Using the comparative approach in the humanities is time-honoured and it is conducive to revealing the unique areas of a group or population by identifying its dissimilarities with a theoretically normative group. In L2 writing research, comparing learner writing with that produced by native speakers or L2 writers from greater proficiency or literacy

groups can help identify features of L2 writing that need to be pedagogically addressed. In the study of syntactic complexity, the researcher employs syntactic complexity measures to compare the performance of two or more writer groups. Alternatively, certain individual grammatical features that are considered to represent mature and complex writing can also be compared.

The emergence of corpus linguistics techniques has given considerable impetus to comparative studies of student writing, since corpus linguistics is inherently favourable for reliable quantitative comparisons. Researchers can make either corpus-based or corpus-driven comparisons between two or more corpora. For corpus-based research, a predefined set of lexis or structures is subjected to frequency count and statistical comparisons. For corpus-driven studies, comparison is derived not from a predefined set but from observing frequency lists, based on which the researcher can further explore the phraseological patterns of high-frequency words and compare their frequencies and meaning-making features in the corpora under study. Part-of-speech (POS) tagging is common in corpus-based research while corpus-driven researchers prefer untagged data with the belief that it is only the observation of raw linguistic evidence that generates new understandings and theories of language. Since the study of syntactic complexity is based on pre-existing, albeit ever-changing, conceptions of the linguistic features of complex writing, the use of corpus data to conduct research in this area should be seen as corpus-based, with POS-tagged and syntactically parsed texts being able to facilitate automatic and semi-automatic search and analysis. Therefore, the present study is a corpus-based one. It is the total accountability to data (McEnery & Hardie, 2012, pp. 14-16) and rapid and reliable access to analysis-related data that makes corpus-based contrastive analysis a valuable methodology.

Both quantitative and qualitative comparisons can be made in contrastive corpus-based analysis of student writing. Quantitative comparisons between corpora can be conducted either with the individual text as the unit of analysis or with the whole corpus as the unit of analysis. In the former case, inferential statistical tests such as the *t*-test or ANOVA can show whether there are significant differences in syntactic complexity measures between two or more corpora. However, data coding and analysis can be time-consuming and laborious for this type of comparison, unless automatic analysis is performed (e.g. Lu, 2011). In the latter, statistical tests like chi-square test and log-likelihood test are usually conducted to compare if a linguistic feature's frequencies in two corpora are significantly different. If affirmative, the feature would either be a case of overuse or underuse in one corpus relative to the other. In addition, comparison can also be made in proportional terms in quantitative corpus-based studies (e.g. Hyland, 1999).

On the other hand, qualitative analysis is conducted by examining the concordance lines of corpus-based evidence (e.g. concordance search) or corpus-driven evidence (e.g. automatically generated *n*-grams). Several types of qualitative comparisons are possible, including the examination of students' misuse of particular linguistic features (e.g. collocations), discourse features of certain grammatical class (e.g. conjunctions), and semi-automatic extraction of certain phrase types (e.g. preposition phrases headed by a particular preposition).

As introduced in 1.6, this study seeks to answer three overarching research questions:

> **Research Question 1**: Which syntactic complexity measures show significant differences between advanced EFL academic writing and expert academic writing?

Research Question 2: What are the grammatical features of noun phrase complexity where EFL student writers lag behind expert writers?

Research Question 3: What are the grammatical features of clause combination where EFL student writers lag behind expert writers?

The study examined the three questions taking a contrastive corpus-based approach. Two corpora were built for this purpose: one of advanced EFL academic writing and one of expert academic writing. The answer to the first question is based on analyses using an automatic computational tool (Lu, 2010) to calculate the syntactic complexity measures for each corpus, with each individual text as the unit of analysis (e.g. Lu, 2011). The second and third questions were examined by careful semi-automatic and manual analyses of various dimensions of noun phrase complexity and clausal complexity; for these analyses, the unit of analysis was the whole corpus. Detailed descriptions of the methodology follow.

3.3. Data

For each of the two groups of academic writing a corpus was compiled: the MA (master of arts) dissertation corpus (hereafter called MDC) and the published research article corpus (hereafter called PRAC).[22] The MDC includes 70 dissertations by Chinese MA students of applied linguistics at seven universities on the Chinese mainland, with a total size of 1,019,243 running words (tokens). That the discipline applied linguistics was selected was mainly due to data accessibility: it was one of the few academic disciplines for which students studying in Chinese universities wrote their MA dissertations in English. Six of the seven universities are institutions specifically devoted to studies of foreign languages and cultures, which are also among

[22] The data used for this study has been previously used in Liu and Li (2016).

the best universities to study English applied linguistics in the country. (See Table 3.1. for the list of seven universities and number of dissertations selected from each). The sample of the 70 dissertations is considered to come from a largely homogeneous population as all of them have been collected in the China Masters' Dissertations Full-text Database, which claims that "it collects distinction-rated dissertations from more than 621 master degree grantors" (CNKI, 2013, my own translation). All the dissertations have been submitted during the past decade or so, representing contemporary EFL academic writing.

The PRAC comprises 129 empirical research articles taken from nine prestigious academic journals in the broad discipline of applied linguistics (see Table 3.2 for details), with a total size of 971,726 running words. All the articles have been published during the past 15 years, chronologically contemporaneous with the MA dissertations collected. Thus, the two corpora are close in size (i.e. around 1 million words), making their comparisons more reliable and meaningful. It should be noted that the PRAC features an international authorship in terms of the authors' L1, meaning that this research is not intended to compare non-native and native speakers of English but is focused on the gaps between advanced EFL students as novice writers and expert writers in terms of syntactic complexity.

TABLE 3.1

List of the Seven Universities for Compiling the MDC

Universities	# of dissertations collected
Xi'an International Studies University	17
Shanghai International Studies University	15
Guangdong University of Foreign Studies	11
Sichuan International Studies University	9
Beijing International Studies University	7
Tsinghua University	6
Tianjin Foreign Studies University	5

TABLE 3.2

List of Academic Journals for Compiling the PRAC

Journals	# of articles collected
TESOL Quarterly	31
Applied Linguistics	30
Journal of Second Language Writing	18
Journal of English for Academic Purposes	11
English for Specific Purposes	10
Language Learning	10
Foreign Language Annals	9
The Modern Language Journal	7
International Journal of Applied Linguistics	3

Both raw text and POS-tagged versions of the two corpora were used in this study. Since most of the analyses in this study were based on the raw texts, special mention will be made only when the POS-tagged corpora were employed. POS tagging of the corpora was performed with the Stanford Log-linear Part-Of-Speech Tagger, to be introduced in detail below.

3.3.1. Discussion of Data Collection

The choice of MA rather than PhD dissertations to serve as the data for advanced EFL academic writing is mainly due to data size and research purpose. As most of the analyses for this study will be done by hand, it would be reasonable to compile a relatively small corpus for detailed study, and a small corpus is also appropriate for grammatical studies as opposed to lexical research. With such a consideration, a corpus of MA dissertations is preferred over one of PhD dissertations in that a much greater number of student writers from a wider range of institutions can be represented in the data sampled. Another consideration of data size is concerned with the size of individual texts. As said in Section 2.2.1, the length of an MA and PhD dissertation is respectively about 10K to 15K words and as much as 80K words. Thus, MA dissertations are more comparable to research articles in terms of individual text length, meanwhile minimising the impact of space on syntactic features, if there is any. Moreover, examining MA rather than PhD data for syntactic complexity features also has pedagogical significance. This is because studying at the master level marks a crucial watershed in the development of academic literacies, i.e. transitioning from proficiency building at the undergraduate level to research-oriented writing skills. Thus, strong pedagogical intervention at the master level is conducive to a more effective preparation for research writing at the MPhil and PhD levels.

As can be seen from Table 3.2, the data makeup for PRAC is unevenly distributed among the nine academic journals, with most of the articles taken from *TESOL Quarterly* and *Applied Linguistics*. Although being a limitation by violating data representativeness, this is actually a consequence of trying to feature international authorship in the corpus. As these two journal titles include more articles by international scholars than other titles,[23] the dataset ends up being slightly biased towards them. Nevertheless, a corpus thus compiled is still highly representative of expert academic writing in applied linguistics and would serve as an eligible comparison corpus for student writing.

Another feature of the data makeup of PRAC is that it represents both academic journals that encourage an accessible style such as *TESOL Quarterly* (TESOL Quarterly Submission Guidelines, n.d.) and those that are "concerned with fundamental theoretical issues" such as *Language Learning* (Aims and Scope, n.d.). Since academic journals' readership and style expectations may have an effect on authors manipulating lexicogrammatical resources, it is expected PRAC thus compiled could include different types of context-driven syntactic complexity features.

3.4. Instruments

3.4.1. The L2 Syntactic Complexity Analyzer

Lu (2010) designed an automated syntactic complexity analyser based on 14 frequently used syntactic complexity measures found in L2 writing development research. Called the L2 Syntactic Complexity Analyzer (L2SCA), the system is designed for advanced EFL writing although it "can be used on any English text in

[23] Although it is impossible to accurately judge the linguistic and cultural backgrounds of scholars just from their bio information, author's names provide the best possible clues in this regard.

which sentence completeness is not a major problem" (Lu, personal communication, 26 June, 2013). The analyser includes five types of syntactic complexity measures: length of production unit, sentence complexity, subordination, coordination, and particular structures (see Table 3.3 for a list and brief description of the 14 measures). In this study, the L2SCA was used to automatically compute fourteen measures of syntactic complexity for each text in the two corpora. Comparing syntactic complexity measures between two or more corpora using the L2SCA requires the use of inferential statistical tests such as *t*-test and ANOVA, with the unit of analysis being the individual text rather than the whole corpus. It also needs to be noted that Bonferroni correction is required for such tests due to multiple comparisons with the same sets of data.

In the automated analysis, the L2SCA first asks the Stanford parser (Klein & Manning, 2003) to analyse the target text(s) into separate clauses, which then are tokenised and tagged for part of speech (POS) with the very parser. This process outputs sequences of parse trees, which the L2SCA automatically calls Tregex (Levy & Andrew, 2006) to query to identify and count the syntactic structures and production units based on manually defined Tregex patterns. Lastly, the analyser computes the 14 complexity measures based on the results of Tregex analyses.

TABLE 3.3

List of Syntactic Complexity Measures Used in the L2SCA

Measure	Code	Definition
Type 1: Length of production unit		
Mean length of clause	MLC	# of words / # of clauses
Mean length of sentence	MLS	# of words / # of sentences
Mean length of T-unit	MLT	# of words / # of T-units
Type 2*: *Sentence complexity		
Sentence complexity ratio	C/S	# of clauses / # of sentences
Type 3*: *Subordination		
T-unit complexity ratio	C/T	# of clauses / # of T-units
Complex T-unit ratio	CT/T	# of complex T-units / # of T-units
Dependent clause ratio	DC/C	# of dependent clauses / # of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses / # of T-units
Type 4*: *Coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
Sentence coordination ratio	T/S	# of T-units / # of sentences
Type 5: Particular structures		
Complex nominals per clause	CN/C	# of complex nominals / # of clauses
Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units

Note. This table is largely based on Lu (2010, p. 479).

Lu (2010) found a high degree of annotator-system reliability for retrieving the syntactic structures and production units related to the 14 measures, with *F*-scores on his test data ranging from .830 for complex nominals to 1.000 for sentences.[24] The system can analyse individual texts or multiple text files in a folder. Results generated by the L2SCA are automatically saved as plain text files and can be easily exported to SPSS or Excel for statistical analysis. A step-by-step manual for using the L2SCA is available in Lu (2014).

The analyser only generates meaningful results when syntactic well-formedness of the input texts is not a major problem, that is, they should be generally free of grammatical errors. It may be true that grammatical errors are common in second language writing at the beginner and intermediate level and they would therefore affect parsing results. But the major linguistic problem with advanced EFL writing at the postgraduate level lies not so much in grammatical well-formedness as in effectively manipulating pragmatics and idiomaticity (Lu, 2010). Therefore, the automatic system is expected to produce accurate parsing results for MA dissertations in terms of identifying the grammatical functions of various constituents.

As discussed in 2.1.3, it needs to be emphasised that the notion of "clause" defined for the measures included in the L2SCA follows the definitional tradition in composition research (e.g. Hunt, 1965), referring only to finite clauses and regarding non-finite structures (e.g. infinitives and participial structures) as part of the finite clause. Therefore, results obtained using the L2SCA should be interpreted differently from studies using clausal indices which also acknowledge non-finite structures (e.g. Biber, Gray, & Poonpon, 2011). For example, measures adopting the broader concept

[24] It needs to be noted that the data with which the annotator-system reliability was calculated was undergraduate essays written by Chinese students, which are relatively grammatically simple. This may result in more structures being accurately recognised by the analyser than when more complicated genres such as MA dissertations are to be processed.

of clause would return a higher clause density in the overall discourse than measures only acknowledging finite clauses. As shown by Yang, Lu, and Weigle (2014), a non-finite clause ratio such as non-finite structures per (finite) clause can be directly derived from the results generated by the L2SCA, obtainable by subtracting 1 from the measure of verb phrases per clause (VP/C), because one finite clause contains one finite VP and the rest VPs are non-finite.

Another conceptual issue that needs to be noted is the definition of the "complex nominals" as used in the measures CN/C and CN/T listed under Type 5 (particular structures) in Table 3.3. Lu (2010) defines complex nominals as including three structural types: 1) NPs with premodification and/or postmodification, 2) nominal clauses (aka. complement clauses), and 3) gerunds and infinitives in subject position. Thus, the two measures of complex nominals examine syntactic complexity of a nominal attribute at both the phrasal (Type 1) and clausal (Type 2 and 3) levels. A drawback of these tripartite measures is that results obtained from them can only be used to explain the overall performance of the three structural types, unable to account for the performance of any individual structural type. Pending demonstration of a strong correlation among the three structural types, it would be necessary for each of the three nominal structures to be separately examined. Especially necessary is that measures be designed for complex noun phrases (i.e. the first component of complex nominals listed above) as this structural type stands out as the most prominent grammatical feature of academic discourse (Biber et al., 2011; Halliday & Martin, 1993). Indeed, Yang et al. (2014) specifically developed a measure of complex noun phrases (i.e. complex noun phrases per clause) through a modification of the Tregex pattern used to identify complex nominals in the L2SCA. However, although this measure can calculate the ratio of the use of complex noun phrases, it still cannot tease out the embedding within

the noun phrase (Biber et al., 2011; Givón, 2009; Givón & Shibatani, 2006), thus calling for further research to address this issue.

In an example application of the L2SCA, Lu (2010) employed the system to tap into a sample from the Written English Corpus of Chinese Learners (Wen, Wang, & Liang, 2005) to evaluate the capacity of the 14 measures to gauge overall proficiency development across three curricular levels. It was found that six of the 14 measures were able to significantly differentiate between-level development: MLC (mean length of clause), MLT (mean length of T-unit), CP/C (coordinate phrases per clause), CP/T (coordinate phrases per T-unit), CN/C (complex nominals per clause), and CN/T (complex nominals per T-unit). In a follow-up large-scale corpus-based evaluation of the 14 syntactic complexity measures to index college-level EFL writing development, Lu (2011) used his L2SCA to analyse 3,554 timed and untimed essays from the same EFL writing corpus, 83% of which belong to the argumentative text type. Among other findings, it was found that good measures in terms of the ability to differentiate between-level development included all the six candidates identified in Lu (2010) and one more length measure, MLS (mean length of sentence).

The L2SCA was designed to track L2 writers' syntactic complexity from a *developmental* perspective. The primary purpose of the system is to tease out those measures that can successfully differentiate between proficiency levels, as in Lu (2011). However, it can also serve as an instrument to profile the syntactic complexity of individual writers or groups of writers, the results of which can be used for comparative analysis, as in Lu and Ai (2015).

3.4.2. Wordsmith Tools 5.0

Wordsmith Tools is a corpus analysis software package developed by Mike Scott (2008), comprising three major tools—Concord, KeyWords, and WordList, in addition to an array of other text management and analysis utilities. This study mainly employed the software's Concord tool to search for particular words and strings for detailed manual analysis of noun phrase complexity and clause combination. A major advantage of the concordancer offered in Wordsmith Tools over other corpus analysis programmes such as AntConc (Anthony, 2014) and ConcGram (Greaves, 2009) is that it easily allows multiple searches to be performed in separate windows, making easier the comparison of search results. However, due to its less developed ability to directly process POS-tagged texts formatted in the "token_tag" sequence generated by the Stanford POS Tagger (to be introduced later), the Wordsmith concordancer was only used to interrogate the raw corpora, while POS-tagged corpora were processed by AntConc, to be introduced directly below.

3.4.3. AntConc

A freely available corpus analysis programme developed and regularly updated by Laurence Anthony, AntConc 3.4.3 (Anthony, 2014) was used in this study mainly for its easy processing of POS-tagged corpora. The search syntax for querying tagged data was based on wild-card search instead of the more advanced regular expression syntax. As an example, a search for the lemma BOOK tagged as a verb by the Stanford POS Tagger can use the input *book*_VB**, where * represents any character and the underscore "_" separates the search word and the designated tag. In addition, AntConc allows the user to show or hide the tags in the concordance window.

3.4.4. Stanford POS Tagger

The POS tagging of corpora in this study was conducted by the freely available Stanford Log-linear Part-Of-Speech Tagger (Toutanova, Klein, Manning, & Singer, 2003), shorted as the Stanford Tagger. Since the default download version of the programme can only tag one piece of text per operation, the task of tagging a folder containing hundreds of text files would be a daunting one. Fortunately, this problem could be easily solved by using a loop script written by Lu (2014) to automatically annotate a folder of text files.[25] The Stanford Tagger (for tagging English texts) uses the Penn Treebank Tag-set, 36 tags in total for coding words and numerals. Although this tag-set would be much smaller by the standard of Lancaster University's 100 strong CLAWS (Constituent Likelihood Automatic Word-tagging System) tag-set, it still satisfied the purpose of this study considering that not much was needed of the grammatical and semantic nuances between different forms of the same lemma for the analysis of syntactic complexity.

3.4.5. SPSS 20.0

SPSS 20.0 (IBM Corp., 2011) was used in this study to perform *t*-tests to compare the means of the 14 syntactic complexity measures automatically computed by the L2SCA. As mentioned above, the results of L2SCA analysis, generated in a text file, can be easily imported to SPSS. Since comparing the means of 14 measures on the same dataset requires multiple comparisons, Bonferroni correction was performed for the *t*-tests (see 3.5.1 for details).

[25] See Lu (2014) for detailed procedure to tag a folder.

3.4.6. Log-likelihood and Chi-square Calculator

The Log-likelihood and Chi-square Calculator 1.0 (Liang, n. d.) is designed used to compute the chi-square or log-likelihood values when comparing the frequencies of a word or linguistic feature between two corpora and to test if the difference was statistically significant. In this study, the log-likelihood test was preferred over the chi-square test for testing statistical significance as a more favoured test used by corpus linguists (McEneary & Hardie, 2012, p. 52). The significance level for frequency comparisons was set at p <.001 in this study, because in quantitative corpus linguistics "[w]hile it is generally accepted that any level below .05 indicates statistical significance, it is quite common for more stringent significance levels to be employed" (Meyer, 2002). Operating the calculator is easy and straightforward, only requiring the user to key in the sizes of the two corpora and the raw frequency in each.

3.4.7. Cohen's **d** *Calculator*

Cohen's *d* is the effect size coefficient indicating the strength of the difference between two means, which has been increasingly required in reporting statistical mean differences in applied linguistics research (Plonsky & Oswald, 2014). Since neither SPSS nor the above two calculators provide Cohen's *d*, it has to be calculated separately, which needs the means and standard deviations in two groups for each variable. With Cohen's *d* ranging from 0 to 1, general cut-off guidelines for its interpretation are: small (0.2), medium (0.5), and large (0.8). Cohen's *d* was computed by an online calculator for effect size available at http://www.uccs.edu/~lbecker/.

3.5. Data Analysis Procedure

The three overarching research questions of this study will be addressed in an individual chapter each (i.e. Chapter 4 to 6). Before presenting the analytical procedure in detail, Table 3.4 previews the analyses to be performed for each research question, and hence each chapter.

TABLE 3.4

Preview of Analyses Performed for Each Research Question

Research Question 1	Chapter 4 – Automatic Measurement of Syntactic Complexity	• Computation of 14 measures of the L2 Syntactic Complexity Analyzer • Comparison of mean scores to test for significant difference

Research Question 2	Chapter 5 – Examining Noun Phrase Complexity	• Complexity of Structural Types of Postmodification
		  ○ Proportion of noun phrases with clausal postmodification
		  ○ Proportion of noun phrases with phrasal postmodification
		• Complexity of postmodification in CNPs
		  ○ Length of postmodification (defined by number of words)
		  ○ Depth of postmodification (defined by levels of postmodification)
		• CNPs as clause subjects
		  ○ Proportion of CNPs as subjects
		  ○ Lexical richness and valency pattern of predicate vocabulary for CNP subjects

Research Question 3	Chapter 6 – Examining Clause Combination	• Frequencies of finite adverbial clauses denoting reason, condition, and concession • Textual analysis of students' use of finite adverbial clauses • Frequencies of participle adverbial clauses • Textual analysis of students' use of participle adverbial clauses

3.5.1. Measuring Syntactic Complexity

This study employed the L2 Syntactic Complexity Analyzer to automatically calculate the values of 14 measures of syntactic complexity for each text in the two corpora. Then, an independent samples *t*-test was run for each measure using SPSS 20.0 to see if there was a significant difference in each measure between the two corpora.

Since 14 measures were to be investigated by 14 *t*-tests on the same dataset simultaneously, the Bonferroni correction was employed to avoid spurious positives. This set the alpha value for each comparison to .05/14, or .004, where .05 was the significance level for the complete set of tests, and 14 was the number of individual tests being performed (cf. Ai & Lu, 2013; Yang et al., 2015).

3.5.2. Examining Noun Phrase Complexity

After examining the 14 syntactic complexity measures, the study went on to examine noun phrase complexity by carefully analysing students' use of complex noun phrases (CNPs), recognised as the most distinctive grammatical feature of academic discourse. Although Biber et al. (1999) have a whole chapter on "complex noun phrases" (i.e. Chapter 8), they do not provide a clear definition of the term. Yang, Lu, and Weigle (2015) designed a CNP ratio measure by modifying the computational pattern for identifying complex nominals in the L2SCA—number of CNPs per clause (CNP/C), where CNPs were defined as NPs that contain one or more noun modifiers: premodifying adjectives, postmodifying preposition phrases, and postmodifying appositives. Although lacking an obvious completeness of NP modifying structural types (e.g. premodifying nouns, and relative clauses), this definition is analytically convenient. Therefore, based on Yang et al., this study defines complex noun phrases as noun phrases with one or more premodifiers or postmodifiers. What are counted as premodifiers and postmodifiers in this study will be introduced in detail later.

The examination of NP complexity in student writing was carried out from four perspectives: 1) the structural types used as NP postmodifiers; 2) syntactic complexity in NP postmodification; 3) the extent to which CNPs serve as grammatical subjects of

clauses, and 4) semantic complexity of CNP subjects through the lens of their predicate vocabulary.

Due to the ubiquity of noun phrases in academic writing, the examination of NP complexity was conducted on a carefully selected set of NP samples. This is because it would be difficult to computationally capture noun phrases in their complete boundaries, especially problematic of which is the identification of preposition phrases as NP postmodifiers at the end of multiple postmodification (Biber & Gray, 2011; Lu, personal communication, 2015). To arrive at an accurate estimate of the use of preposition phrases as NP postmodifiers, Biber and Gray (2011) used a randomly selected subsample of postmodifying preposition phrases from their corpora by coding "every fourth occurrence of the prepositions *in*, *on*, *with* and *for* that was not preceded by a word tagged as a verb" (p. 227). However, the present study did not use random selection as did Biber and Gray, but selected a subsample of noun phrases headed by a number of carefully selected high-frequency academic nouns. This more focused sampling method allows for an in-depth investigation into students' production of NPs from various perspectives vis-à-vis expert use. In addition, comparing NPs with the same set of head nouns in the two corpora not only allows the researcher to make overall comparisons of noun phrase complexity but also to assess students' performance on NPs with each individual head noun. Since patterns of postmodification may vary for different head nouns as determined by their semantic attributes—for example postmodification patterns for *difference* and *implication*—the results obtained from only observing an NP subsample as a whole might not provide pedagogically informative implications for building NP complexity.

NP sampling procedures are described below, followed by the detailing of the analytical procedures for each of the four aspects to noun phrase complexity outlined above.

3.5.2.1. NP Sampling Procedures for Examining NP Complexity

The examination of NP complexity in this study was based on an NP subsample headed by a set of six nouns: *development*, *information*, *relationship*, *difference*, *analysis*, and *knowledge*. The six head nouns were arrived at by a set of selection criteria. First, it was decided that the potential head nouns should be abstract nominalisations, i.e. nouns derived from verbs, adjectives, and occasionally other word classes. The main reason nominalisations other than common nouns were considered is that they are semantically abstract and therefore usually need modification or complementation to substantiate their meaning; hence greater possibility for NP complexity to develop from nominalised head words. Second, the head nouns should be among the top 10 nominalisations in Gardner and Davies' (2013) "Academic Vocabulary List" generated from the 120 million word academic subcorpus of COCA (Corpus of Contemporary American English).[26] This newly developed word list was preferred for this study over Coxhead's (2000) more widely known list because it lists words in their lemmas instead of word families and because it ranks the words in decreasing frequency order rather than alphabetically; this makes words of different parts of speech directly discernible and retrievable by virtue of their frequency information. These features met the need of the present study to locate high-frequency nouns in academic discourse as potential NP head nouns. Third, the nominalisations should have a frequency of over 500

[26] This study used the List's 3,000 most frequent core academic words (lemmas), downloaded at http://www.academicwords.info/download.asp upon request.

occurrences per million words in the two corpora as an arbitrary cut-off point for high frequency nouns.

These three criteria have yielded six nouns as listed in Table 3.5, together with their respective frequencies in COCA-Academic, the PRAC, and the MDC, normalised to # per million words (see McEnery & Hardie, 2012, pp. 49-50 for details on frequency normalisation); the four nouns that met the first two criteria but not the third one are also listed with a strikethrough indicating their exclusion from analysis. This is how the six NP head nouns were arrived at. In terms of word formation, four of the six nominalisations are derived from verbs (*development*, *information*, *analysis*, and *knowledge*), one from a noun (*relationship*), and one from an adjective (*difference*).

TABLE 3.5

Top 10 Nominalisations in Gardner and Davies' (2013) Academic Vocabulary List

		COCA-Acad	PRAC	MDC
1	*development*	529	762	851
2	*information*	516	826	1582
3	~~*activity*~~	460	263	279
4	*relationship*	423	516	545
5	*difference*	409	1096	1224
6	*analysis*	401	1188	1214
7	*knowledge*	305	1421	2165
8	~~*performance*~~	303	800	467
9	~~*organisation*~~	295	205	152
10	~~*movement*~~	244	68	46

Note. The frequencies of the words were normalised to occurrences per million words.

It should be pointed out that for most of the analyses of NP complexity the NPs based on these six head nouns included both independent discrete NPs and those

contained within a higher-level NP. For example, the NP headed by *development* is independent and complete in *the development of web technologies and text analysis software* but is embedded in *useful tools* ***in the development of language-specific level descriptions***. Therefore, the examination of multiple postmodification in this study was performed on complete NPs for reliable inclusion of all postmodifiers for a particular head noun.

Now that the scope of noun phrases for the study of NP complexity has been clarified, the following sections will present the analytical procedures for the four aspects of NP complexity outlined above.

3.5.2.2. Identifying Postmodifiers

In this study only postmodifiers with a restrictive function are considered as forming part of the complex noun phrase (CNP), with non-restrictive postmodification which provides additional, descriptive information on the head noun excluded from analysis. Typographically, while non-restrictive postmodification is usually separated from the head noun by a comma or dash, there is no punctuation at work for restrictive postmodification. Within the restrictive scope of postmodifiers, two major structural types can be distinguished: clausal postmodifiers and phrasal postmodifiers. Clausal postmodifiers include two finite clause types, i.e. relative clauses and complement clauses,[27] and three non-finite types, i.e. present participle clauses (shortened as *-ing* clauses), past participle clauses (shortened as *-ed* participle clauses), and infinitive clauses (shortened as *to*-clauses). Phrasal postmodification with a restrictive function mainly includes preposition phrases and less frequently adverb phrases, adjective

[27] Although noun complement clauses are usually regarded as performing complementation rather than postmodification, they have been included in this study as a type of postmodifier due to their ability to elaborate on the propositional content signalled by the head nouns.

phrases, and noun phrases. The taxonomy of structural types counted as postmodifiers in this study is presented in Table 3.6 with examples featuring each grammatical structure.

TABLE 3.6

Postmodifiers for the Analysis of NP Complexity

Postmodification type	Grammatical structure	Examples from PRAC
Clausal postmodification	relative clause	information ***that has been forgotten***
	complement clause	the knowledge ***that doing so would improve student performance***
	-ing clause	further analysis ***estimating school means***
	-ed clause	identity development ***associated with language learning***
	to-clause	sufficient skills and knowledge ***to accomplish his job***
Phrasal postmodification	preposition phrase	knowledge ***of collocation***
	adverb phrase	a possible way ***out***
	adjective phrase	results ***faithful to Adair-Hauck et al.'s (2010) dialogic***…

Note. Postmodifiers are marked in bold.

The identification of the postmodification status of the above listed grammatical structures was manually vetted to ensure that each instance of the structures in the corpora was a true case of the postmodifier of a particular head noun. For example, in the extended complex noun phrase *the methods of discourse analysis outlined by Fairclough (2003) as well as by Bucholtz and Hall (2005)*, the *-ed* clause *outlined by Fairclough (2003) as well as by Bucholtz and Hall (2005)* following the targeted NP

head *analysis* does not actually postmodify *analysis* but rather the previous noun *methods*. The manual vetting process has thus barred the inclusion of such gapped postmodification from the analysis.

3.5.2.3. Nouns with Postmodification in General

The study of noun phrase complexity started with examining the general tendency of using postmodification with the six head nouns in the two corpora. This tendency was determined in proportional terms, by calculating the proportion of NPs with postmodification to the total number of NPs with a particular head noun. Attention was paid to two aspects to the use of postmodification: the extent to which student writers tend to postmodify nouns, and the tendency of different head nouns to take postmodification. The latter focus of attention makes sense in that nouns that have different semantic and lexicogrammatical properties may differ in their tendency to be postmodified.

3.5.2.4. Complexity of Structural Types of Postmodification

The examination of postmodifying structural types was based on Biber et al.'s (2011, pp. 30-31) hypothesised developmental stages of grammatical structures. The grammatical features used to build noun phrases from this taxonomy were employed by Parkinson and Musgrave (2014) to measure noun phrase complexity of EAP students' academic writing. This study utilised part of their framework to investigate the distribution of structural types serving as noun postmodifiers in the MA dissertation corpus. As noted above, postmodifiers in this study are classified into two major structural types: clausal postmodifiers and phrasal postmodifiers. Table 3.7 re-presents all the grammatical structures from the two broad types, together with their

hypothesised stages and examples. It can be seen that clausal postmodification has been hypothesised to develop at Stage 3 and 4 while part of phrasal postmodification would likely develop at a more advanced stage, i.e. Stage 3 and 5. On another note, non-finite clausal postmodifiers (*-ing* and *-ed* clauses) would develop subsequent to their finite counterparts, whereas postmodifying preposition phrases (PPs) headed by prepositions other than *of* would occur after *of*-PPs. These two predictions have been supported by Parkinson and Musgrave (2014, p. 54), who showed that the use of participial structures and PPs with prepositions other than *of* expressing abstract rather than concrete/locative meanings increased across curriculum levels. However, this study did not distinguish between concrete/locative and abstract PPs as did Biber et al. (2011) and Parkinson and Musgrave (2014), considering that academic discourse is fundamentally abstract.

TABLE 3.7

List of Postmodifying Structures Included for Analysis

Structural classification	Grammatical structures	Sub-structures	Examples from the PRAC	Stage (Biber et al., 2011)
Clausal postmodifiers	Finite restrictive relative clauses	*that*-clauses	*a **difference** that could prove to be quite important*	3
		which-clauses	***knowledge** which "belongs to" others*	3
	Non-finite restrictive relative clauses	*-ing* clauses	*reliability **analysis** indicating a Cronbach's alpha coefficient of .95;*	4
		-ed clauses	***differences** associated with variations in the quality of schools*	4
		to-do clauses	*the **information** to follow*	[a]
Phrasal postmodifiers	Prepositional phrases	*of* phrases	*his **understanding** of its meaning*	3 & 5
		Phrases other than *of*	*summarized or translated source **information** across disciplines*	3 & 5
	Adjective and adverb phrases	Adjective phrases	*the tonal **information** independent of the syllable segmental information*	[a]
		Adverb phrases	*conceptual **knowledge** alone*	[a]

[a] Not included in Biber et al. (2011).

By examining the postmodification pattern of a sample of noun phrases headed by six abstract nouns, the study intended to determine the extent to which student writers use postmodifiers of different structural types in ways similar to and different from expert writers, both in terms of structural distribution for the NP sample as a whole and for NPs expanded from each individual head noun. Three criteria were used to gauge the syntactic complexity of postmodifying structural types: 1) variety of the structural types used, 2) greater use of phrasal postmodifiers vis-a-vis clausal ones, and 3) greater use of non-finite vis-a-vis finite clauses. Although preposition phrases no doubt constitute the major type of postmodifiers in academic writing, using different types of grammatical structures reflects more flexible and resourceful means of packing

meaning and semantic relations in postmodification. The second criterion is based on the premise that more advanced writing would show greater representation of phrasal rather than clausal postmodification than less advanced writing. However, it needs to be noted that more advanced writers tend to use more participial postmodifiers than novice writers, forming the basis of the third criterion. This chapter looks to see how student writers differ from expert writers in the complexity of postmodifying structures thus conceptualised.

Different from previous research examining postmodification, this study is based on a sample of noun phrases expanded from high-frequency abstract nominalisations in English academic texts. While incomplete in the scope of NP coverage, such an approach allows for comparison of postmodification patterns for the NP sample as a whole and for NPs with each individual head noun, considering that postmodification patterns vary across nouns which differ in their semantics and therefore colligational patterns. Thus, more specific information on the postmodification patterns for different head nouns can be obtained, some of which could be at odds with the overall profile of structural types as presented by Parkinson and Musgrave (2014). This approach can also provide pedagogically useful data-driven implications (Warren, 2016) for postmodification resources for particular (high-frequency) nouns. It should be regarded as an important complementarity to generalisations about major postmodifying structures in academic discourse, marking the significance of observing local lexicogrammatical environments of particular nouns to study such unique syntactic properties as postmodification.

3.5.2.5. Syntactic Complexity in CNP Postmodification

In the 5-stage development hypothesised by Biber et al. (2011), extended embedding in the complex noun phrase with consecutive prepositional postmodifiers is considered to be the most complex feature. In the present study, multiple postmodification was not limited to consecutive prepositional postmodifiers, but also incorporated other structural types, for example:

(1) *two L2 French learners' internalization **of** corrective feedback **received** from a native speaker **on** a story [**that**] the two students had written together*

where postmodification includes two prepositional phrases and two relative clauses. Also, the study did not distinguish between concatenation and nesting (i.e. embedding) in multiple postmodification (Oostdijk & Aarts, 1997), regarding both types of postmodification as ways of information expansion on the initial head noun.

Since the three NP complexity measures in the L2SCA do not directly tap into the complexity of multiple postmodification, it can only be examined by careful manual investigation. In order to best analyse multiple postmodification, the data used for this analysis was limited to complex noun phrases whose first postmodifier was a preposition phrase (such as Ex. 1 above), because in English multiple postmodification mostly occurs when the first modifier is a preposition phrase (Biber et al., 1999, p. 642). To further limit the data for examining postmodification to a manageable size, this study only analysed NPs whose first postmodifier is an *of*-phrase, because *of* is the most frequent and representative preposition in English (Biber et al., 1999, pp. 635-642). Finally, NPs built on three head nouns were examined: *understanding*, *analysis*, and *lack*. The selection of these three nouns from which complex NPs are expanded arose from two considerations: 1) they share more or less comparable frequencies in the two corpora when taking an *of*-phrase; 2) all three nouns appear in Gardner and Davies'

(2013) *Academic Vocabulary List*. Including different head nouns permits a broader representation of complex NPs, considering that nouns differ in the complexity of postmodification determined by their internal semantic attributes. For instance, the postmodification for *relationship* and *effect* could be potentially intricate as both nouns can appear in complex phraseological patterns, i.e. *relationship between…and…* and *effect of…on…*:

(2) *a close* ***relationship between*** *general second language anxiety* ***and*** *self-perceived competence in L2*

(3) *the* ***effects of*** *the intensive summer Mandarin program* ***on*** *students' language achievement as measured by the SOPI test*.

As noted above, only independent, complete NPs headed by these three nouns were included, with those contained in a higher-level NP excluded.

Table 3.8 lists the occurrences of NPs headed by these three nouns in the two corpora. Although limited in comprehensiveness, the selected complex NP sample can still meet the objective of analysing and comparing postmodification in the two corpora.

TABLE 3.8

Instances of CNPs Expanded from Three Head Nouns in the Two Corpora

	MDC	PRAC
understanding	232	251
analysis	272	275
lack	152	224

Note. The same set of CNPs have been used in Liu and Li (2016).

Two aspects in the use of postmodification were investigated: 1) length (or number of words) of postmodification and 2) depth of multiple postmodification,

referring to the number of levels of multiple postmodification of the selected head nouns and the proportion of multi-postmodified CNPs at each level. Extraction of the multi-postmodified CNPs was performed with WordSmith's concordance tool but the identification of the complete postmodification and each level of postmodifier was manually determined. Subsequently, the length of the complete postmodification was calculated for each instance of CNP and saved in SPSS for independent *t*-tests of mean length comparison; and the number of postmodification levels for each CNP was counted for comparing postmodification depth.

Two tricky issues in the identification and segmentation of multiple postmodifiers need to be addressed, both of which have been found to cause confusion and indeterminacy in the analytical process. First, when a postmodifier contains a clause that further contains an NP with postmodification, the extra postmodifier(s) within the clause-embedded NP would not be included for the analysis. For example, in the CNP

(4) *knowledge of how to distinguish agent/patient roles **in Latin***,

the preposition phrase *in Latin*, as a postmodifier of the noun *roles* which is itself part of the "*how to*" clause in the *of*-headed preposition phrase postmodifying *knowledge*, was not regarded as an extra level of multiple postmodifier. For another example, in

(5) *the analysis of the moves that authors make when organizing the principal sections **of their RAs***,

the preposition phrase *of their RAs* appears in a *that*-relative clause, therefore not qualified as an extra level of postmodification. Second, when there are two or more coordinate phrases at a particular level of postmodification, the postmodifier (if any) in the last coordinate phrase would be counted as an extra level of postmodification. For example, in

(6) *analysis of the interview data* ***and open-ended questions of the survey***,

the preposition phrase *of the survey* postmodifies *questions* in the PP-embedded NP *open-ended questions* coordinating with the previous PP-embedded NP *the interview data*, thus being qualified as an extra level of postmodification.

3.5.2.6. CNPs as Grammatical Subjects

The fourth and last analysis of NP complexity brought its examination under a wider context by looking at its involvement in the expansion of the clause. Specifically, the analysis examined the extent to which complex noun phrases serve as grammatical subjects in finite clauses. The idea of tapping into this dimension of NP complexity was derived from two sources of motivation: 1) the use of inanimate/abstract subjects is a distinctive grammatical feature in academic writing and 2) it is not uncommon to use highly information-rich inanimate/abstract subjects in the form of heavily postmodified complex noun phrases and other forms of complex nominals (e.g. non-finite clauses and nominal clauses). See 2.4.2.3 for a review of Vande Kopple's (1994) work on this grammatical feature.

However, examining the syntactic configurations of subject-position CNPs is just the first step to learning about the roles they play in the clause and the relationships they have with other elements of the clause, especially the verbal processes in which they are a part, as the complexity of the clause gradually arises. Therefore, the next step would be to examine the CNPs in association with the verbal processes in which they serve as clause subjects and how lexicogrammatical complexity involved in the processes can be explained.

3.5.2.6.1. Counting CNP Subjects

Still utilising the set of CNPs headed by *understanding*, *analysis*, and *lack*, the analysis coded the instances where the CNPs served as grammatical subjects of finite clauses and calculated the proportions of subject-filling CNPs to all CNP instances in the two corpora. Subsequently, the mean lengths of subject-filling CNPs in the two corpora were compared for statistical significance. The analysis further distinguished between "extremely" lengthy subject-filling CNPs and "common" ones by setting up an arbitrary cut-off point of 10 words for differentiating the two length categories, with CNPs ≥ 10 words considered as "extremely" lengthy subjects and those shorter than 10 words as "common". These three aspects of subject-filling CNPs were expected to provide a prismatic view of how student writers use long subjects as a way of organising and foregrounding complex information.

Although CNPs can occur in a range of syntactic positions (Biber et al., 1999, pp. 98-99), the present analysis only examined their occurrences as clause subjects. Such a methodological choice lends itself to the study of the extent to which CNPs fulfil the role of inanimate/abstract participants in the clause, a linguistic feature typical of academic discourse.

3.5.2.7. Predicate Vocabulary with CNP Subjects

The study went on to examine the predicative processes initiated by the CNPs under investigation, i.e. when they serve as clause subjects. It aimed to reveal the potential contribution the CNPs make to the overall clause-internal complexity as the clause unfolds, starting from the subject and throughout the predicate. The previous analysis examining the density and length of CNP subjects does not offer much about the kind of relationships the CNPs enter into with the predicate system of the clause. In

other words, it would be necessary to examine the ways in which the CNPs in question contribute to the expansion of the clause, which requires an analysis of the verbal processes initiated by CNP subjects.

As discussed in 2.4.2.3, inanimate subjects are usually used with active verbs in academic writing, including lexical verbs and copular verbs with various valency patterns. Lexical verbs preferred in this highly elaborated genre are predominantly of a Graeco-Latin origin (Coxhead & Byrd, 2007) and from a wide range of semantic domains (Biber et al., 1999). Apart from being loaded with a formal tone, Graeco-Latin vocabulary in general entails more specific meanings than Anglo-Saxon words and is therefore more appropriate for providing detailed description and in-depth explanation and argumentation integral to academic writing. Of course, this is not to say that Anglo-Saxon verbs are rejected as less valuable and it is the selection of context-pertinent items that matters more.

With respect to the semantic domains of verbs used in academic discourse, Biber et al. (1999, p. 360ff) have shown that activity verbs, existence verbs, and mental verbs occur the most frequently, although verbs from the communication, occurrence, and causative domains are also frequently used. Thus, while inanimate subjects are usually abstract nominals, the verb phrases that go together with them more often than not exhibit features of human engagement, which, however, is carried out by the non-human participants. This is how grammatical metaphor works in its essence. A large number of such human-involving verbs frequently used in the Introduction, Literature Review, and Results sections of research articles and dissertations have been extensively researched under the umbrella term **reporting verbs** (e.g. Hyland, 1999; Thomas & Hawes, 1994; Thompson & Yiyun, 1991), often associated with the examination of authorial stance ingrained in the configuration of subjects and verbs.

Thus, a generalised suggestion that EFL writers use significantly more private verbs (i.e. mental verbs) and public verbs (i.e. communication verbs) than L1 writers such as that provided by Hinkel (2003) needs to be re-evaluated based on the particular verbal items and their valency patterns employed, in view of the fact both these two types of verbs occur frequently in academic prose.

In this study, the analysis of predicate vocabulary with CNP subjects was focused on lexical verbs and predicative adjectives and prepositions used with copulas. The study of complexity in this regard focused on three aspects of the verbs: 1) valency patterns, 2) semantic domains, and 3) lexical richness. With the scope of CNP subjects being certain and limited, a greater variety of these aspects represents a greater scope of meaning assigned to the subjects. It needs to be kept in mind that the examination of the variety and sophistication of the predicate vocabulary with CNP subjects did not arise from an interest in student writers' decontextualised vocabulary knowledge per se, but rather aimed to reveal the scope and complexity of meanings students are able to assign to abstract subjects through the selection of predicate vocabulary. A brief introduction to verb valency, semantic domains, and lexical richness is given below.

In grammars that see the verb as central to clause structure, verb valency refers to the number of arguments (or complements) controlled by a verbal predicate. It is related to, but not identical with, the notion of verb transitivity, which only concerns object arguments of the verb predicate. With most verbs requiring an obligatory subject, it would be more meaningful to examine the patterns following the verb. Therefore, this study adopted traditional grammar's classification of verb valency or transitivity in English (e.g. Biber et al., 1999; Quirk et al., 1985), only examining the patterns following the verb.

Basically, verbs can be divided into copulas and transitive and intransitive lexical verbs according to their ability to directly take complements. While all copular verbs must have complements, the valency of lexical verbs takes on a complicated picture. Transitive verbs can be distinguished by monotransitives taking a single direct object (e.g. ***produce*** *replicable results*, ***delete*** *redundant information*), ditransitives taking an indirect object and a direct object (***give*** *due credit to the role of DMs*, ***put*** *these results on a firmer basis*, ***accuse*** *students of plagiarism*), and complex transitives taking a direct object and an object complement (e.g. ***make*** *the classroom experience more authentic*, ***call*** *it a model*).[28] On the other hand, intransitive verbs cannot take direct objects—they are either used independently (e.g. *this process* ***occurs*** *early in service, a pattern* ***emerged*** *for all participants*) or with a preposition to become what is called prepositional verbs (*reviewers thus* ***complied with*** *the instructor's request*, *he* ***identified with*** *our distress and despair*). It should be noted that examining valency patterns in student writing was intended to see how diversified a scope of predicate vocabulary student writers are able to produce for fulfilling the meaning of abstract CNP subjects. It did not mean to see if students are able to use these verbs accurately.

Lexical richness is an umbrella term referring to different dimensions of lexical complexity, usually including lexical density, lexical variation, and lexical sophistication (Lu, 2012). Lexical density, which refers to the proportion of lexical words in discourse, can be used as a criterion to distinguish distinct registers, such as speech and writing. Since the examination of predicate vocabulary was only concerned with lexical words, lexical density was not relevant here, with only the other two dimensions examined.

[28] Object complement is called object predicative in Biber et al. (1999).

Sometimes also called lexical diversity or lexical range, lexical variation "refers to the range of a learner's vocabulary as displayed in his or her language use" (Lu, 2012, p 192). Lexical sophistication measures "the proportion of relatively unusual or advanced words in the learner's text" (Read, 2000, p. 203). This measure is defined on the basis of word frequency by reference to a general corpus (e.g. the BNC or COCA), with texts containing a higher rate of low-frequency words being more sophisticated.

Given the fact that texts from the two corpora are from the same academic discipline and both fall into the written academic register, i.e. empirical studies in applied linguistics, the scope of predicate vocabulary taken by CNP subjects should not so much be affected by the selection of texts as by the possibilities of event, relation, and action which the authors believe may apply to those inanimate/abstract subjects through the predicate vocabulary. Therefore, lexical variation and sophistication in the choice of predicate vocabulary for CNP subjects in a corpus reflects the breadth and depth of meaning potentials associated with abstract subjects, hence a more sophisticated and varied semantic representation mapped onto the CNPs concerned. Thus, assessing the breadth and depth of meaning assigned to CNP subjects represented in the choice of predicate vocabulary was in order as part of the examination of noun phrase complexity.

3.5.2.7.1. Analysing Valency Patterns and Lexical Richness of Predicate Vocabulary

Specifically, vocabulary items in the predicate vocabulary taken by the selected CNP subjects in the two corpora were manually identified and listed as types (as opposed to tokens) together with their number of occurrences. They were subsequently analysed for their valency patterns and lexical richness.

The following verb valency patterns were established to categorise the predicate vocabulary taken by the CNP subjects. A distinction was made between transitive, intransitive, and copular verbs, with each category comprising different valency patterns to be identified.

TABLE 3.9

Valency Patterns of Predicate Vocabulary of CNP Subjects

Transitive verbs	**Intransitive verbs**	**Copular verbs**
Monotransitive	Used alone	Copula + noun phrase
Ditransitive	Prepositional verbs	Copula + adjective phrase alone
Complex transitive		Copula + adjective phrase + complement
		Copula + preposition phrase

The examination of lexical variation and sophistication of the predicate vocabulary made use of lexical complexity measures developed in L2 vocabulary acquisition research (Lu, 2012). Lu reviewed a number of measures that gauge lexical variation, including number of different words (NDW), type-token ratio (TTR) and its various transformations, with each measure having its own strengths and weaknesses. This study employed the "intuitively straightforward" measure NDW as the basis for examining the variation of the predicate vocabulary taken by CNP subjects. However, there was a slight modification to this measure: instead of counting the types of vocabulary items in their different inflectional forms, the study enumerated their lemmas that have been attested in the corpora: thus, *do*, *does*, *did*, *done*, and *doing* counted as one lemmatised type, i.e. *do*.

With respect to lexical sophistication, researchers can set the frequency cut-off point in accordance with the general literacy level of the students involved in their

studies. For example, Lu (2012) regarded as sophisticated words beyond the 2,000 most frequent items generated from the BNC for his study of undergraduate students' written lexical sophistication. However, a major problem facing the present analysis was that the frequency of multi-word sequences used in the predicate such as prepositional verbs (e.g. *result in*) and complex prepositions (e.g. *is in front of*) cannot be obtained from any frequency-based word lists as these lists only provide information on individual words. If multi-word items are to be considered as semantically non-compositional and functioning similarly as single-word verbs, it would be more appropriate to decide their frequency in their own right as a whole rather than on the basis of their core verbs: *to draw on* does not occur with the same frequency and significance as *to draw*. Thus, to obtain a more accurate frequency profile of predicate vocabulary required that single-word and multi-word verbs be dealt with separately for the calculation of frequency. For this reason, the present study made use of the academic subsection of the Corpus of Contemporary of American English (COCA) to check the frequency of both single- and multi-word items, which could be easily performed with its online interface (http://corpus.byu.edu/coca/). A greater representation of low-frequency items marks a greater degree of lexical sophistication of predicate vocabulary used with CNP subjects.

3.5.2.7.2. Association with Collocation Research

It is necessary to emphasise that the examination of vocabulary choice for predicate vocabulary taken by CNP subjects extends beyond the interest in the use of collocation, in that collocation research is more concerned about examining formulaic sequences consisting of adjacent items (e.g. bi-grams or *n*-grams) rather than the selection of predicate vocabulary occurring much further to the subject. While comparing the variety of predicate vocabulary may fall into the broader research

domain of collocation which is largely concerned with examining the acceptability and idiomaticity of adjacently co-occurring words (e.g. Nesselhauf, 2003), the focus of the current analysis is not on the co-occurrence of adjacent words, nor is it much interested in finding out phraseological errors. Strictly speaking, the analysis here would not equate itself with collocation research involving nouns and verbs in that collocation research on these two word classes mainly focuses on **verb-noun** sequences (e.g. Laufer & Waldman, 2011; Nesselhauf, 2003), rather than the other way around, i.e. noun-verb co-occurrences.

Indeed, the selection of predicate vocabulary for abstract subjects does not so much depend on conventionalised, psycholinguistically-conditioned co-occurrence of lexical items (Ellis & Frey, 2009) as on more general semantic compatibility. For example, while *make a trip* and *make a tour* would sound more familiar and idiomatic than, say, *make a travel*, as constrained by the Idiom Principle (Sinclair, 1991), the very principle may not apply when *make* is selected for its subject; thus, any consciousness-bearing individual is able to *make a trip* or *journey*, e.g. *I*/*My dad*/*My family*/*Our class will make a trip to Hong Kong next week*, where it is more probably the Open-Choice Principle (Sinclair, 1991) that is in operation. Therefore, the current analysis, which aimed to identify the range of predicate vocabulary following the CNP subjects, should be understood as examining the authors' repertoire of semantic possibilities that can be mapped onto the abstract and complex subjects. Arguably, this would be a largely insightful and visionary meaning-making process under the Open-Choice Principle.

This point could be further supported by the fact that entries of the three head nouns used for this analysis, i.e. *understanding*, *analysis*, and *lack* in the Oxford Collocations Dictionary (Lea, 2002) cover little usage on their **noun-verb** collocations: no verb collocates are provided for *understanding* and *lack*, and there are only two

verbs listed for *analysis*—a coverage much less limited than the actual co-occurring instances attested in corpora. As the results will show, even the student corpus exhibits a far wider variety of **noun-verb** instances than the collocation dictionary provides. Since academic writing requires writers to select lexical items that best suit their intentions, such a meaning-making process is not always a restructuring of prefabricated chunks or bundles as would be more reasonably needed in spontaneous speech. In this sense, the cognitive-linguistic complexity and challenge posed by assigning meanings to abstract and complex subjects appear to mainly operate under the Open Choice Principle.

3.5.3. Examining Clause Combination

This study examines clausal complexity as the other major component of syntactic complexity along with its conceptual counterpart noun phrase complexity, whose analytical rationale and procedures have been explicated above. The examination of clausal complexity in this study was focused on a major form of clause combination, i.e. the combination of what is traditionally called adverbial clauses with a main clause. Its identification and analytical procedures are detailed below.

3.5.3.1. Foci of Clause Combination

Two types of clause combination can be distinguished: coordination and subordination, or parataxis and hypotaxis in functional terms. However, since clausal coordination does not have a strong presence in academic writing and is not grammatically complex, the features of clause combination in student writing will be mainly focused on subordination. However, the dependent clause types included in the measures of subordination in the L2SCA do not all have a strictly "clause-combining"

function. Among the three dependent clause types (i.e. adjectival, adverbial, and nominal),[29] only adverbial clauses should be regarded as clause combination, with the other two performing an embedding function (embedded in a higher-level clause or in a noun phrase). Ex. 7 below exemplifies how adverbial clause is combined with the main clause, Ex. 8 shows how adjectival clause is embedded in a noun phrase, and Ex. 9 is an example of nominal clause, with the dependent clauses highlighted in bold:

(7) ***Because these quantitative outcomes have varied greatly with the amount of supportive mediation****, the present work aims to elucidate the relationship between…*

(8) *Second language (L2) theorists and practitioners have long recognized a distinction between the explicit knowledge about language* ***that results from conscious, analytical reflection****, and the implicit, intuitive knowledge of language* ***that underlies unplanned, real-time communication***.

(9) *Example 1 demonstrates* ***how our coding was applied to an excerpt of small-group Co-construction in Lesson 1****….*

Embedding, as in Ex. 8 and 9, should be regarded as having a nominal property and thus cannot be understood as a clause combining device. Therefore, only adverbial clauses are included in the investigation of clause combination in student writing.

As described in 2.1.3, the present study followed the clause grammar of English to regard non-finite structures as belonging to the category of clause rather than verb phrase, hence their inclusion for the study of clause combination. Moreover, the distinction between embedding and hypotaxis for finite dependent clauses also holds for non-finite clauses. Similar to finite clauses, non-finite clauses having adjectival (e.g.

[29] Adjectival clause and nominal clause called in the L2SCA are also known as restrictive relative clause and complement clause respectively.

those serving as NP postmodifiers) and nominal (e.g. those serving as subjects or verb complements) properties are embedded structures and adverbial non-finite structures are clause-combining devices. These three types of participle clauses are respectively exemplified below, again with the dependent participle clauses highlighted in bold:

(10) *The conflicting outcomes of studies* ***comparing deductive and inductive L2 instruction***[participle clause as postmodifier] *suggest the need to explore a balance between the two extremes*.

(11) ***Developing a firmer empirical grasp of how it works in the reality of the classroom***[participle clause as subject]*, then, is pivotal to educators' understanding of teacher practices and, ultimately, to their ability to implement effective teacher training*.

(12) *In this article, I use understanding-check in its narrow sense****, referring to teachers' use of yes-no questions to seek confirmation on whether the just-prior activity or talk has been received without any problems***[adverbial participle clause].

In this study, only participle adverbial clauses were included for the examination of clause combination considering their prominence in the written registers and EFL learners' underdeveloped sensitivity to participle clauses in their essay writing (Granger, 1997).

In what follows, some of the L2SCA measures involving subordination will be decomposed to address different clause combining processes. The clause types that will be teased apart for detailed examination include circumstance adverbial clauses (structures included in the dependent clause measures) and non-finite participle adverbial clauses (structures included in the verb phrase measures).

3.5.3.1.1. Finite Adverbial Clauses

The analysis of clause combination first focuses on one of the three components of dependent clauses used in the L2SCA—the finite adverbial clause. Although found to be significantly more frequent in informal conversation than in academic writing and thus argued to be less representative of the grammatical complexity of academic prose (Biber et al., 2011), finite adverbial clauses are actually important meaning-making devices in establishing various logical relationships between propositions, claims, and arguments.

The semantic force (Schleppegrell, 1996) encoded between the adverbial and the matrix clause in academic writing is usually much greater than in all other registers, considering the already enormous clause-internal noun phrase complexity in clause elements. Common semantic categories of finite adverbial clauses in academic writing include reason, concession, and condition, crucially contributing to building coherent and complex discourse semantics of the unfolding exposition and argumentation. On the one hand, reason adverbial clauses are used to provide cause, support, explanation, or justification for a particular statement, a methodological action taken, or a result obtained. On the other hand, both conditional and concessive clauses are integral to the presentation of arguments: while "conditionals are often used to introduce and develop arguments.…concessive clauses are often used to show the limitations of certain facts, events, or claims" (Biber et al., 1999, pp. 824-825).

3.5.3.1.2. Participle Adverbial Clauses

An important feature of advanced academic writing is the use of reduced, phrasal structures to incorporate more information into a finite clause. Two major processes can be employed to convert a finite clause to phrase-like structures: reducing

a finite clause to a noun phrase or preposition phrase, or reducing a finite clause to its non-finite or verbless counterpart. A salient function of non-finite structures in academic writing as opposed to conversation concerns the use of present and past participle structures (also referred to as "-ing" and "-ed" participles) to perform a couple of grammatical functions due to their versatile syntactic property, i.e. they are partly nominal (only for *-ing* clauses), partly adjectival, and partly verbal. Being skilful at manoeuvring non-finite structures not only increases the variety of syntactic structures but more importantly it is capable of contributing to the development of local coherence of the evolving discourse. More often taken as non-finite forms of verb phrases in L2 writing research, participle structures are usually analysed as non-finite clauses in grammar and discourse studies. This study adopts the view that participle structures should be given a clausal status, albeit a reduced and incomplete form of clause. Therefore, the use of participle structures is seen as a case of clause combination rather than phrase combination.

Although measures of verb phrase ratios in the L2SCA (e.g. verb phrases per T-unit) partly cover non-finite verb structures, they cannot adequately explain the performance of non-finite clauses since it automatically captures both finite and non-finite verbs (Lu, 2010, pp. 481-484). A roundabout way to calculate the ratio of non-finite verbs per (finite) clause was employed in Yang et al. (2015) who subtracted 1 from the number of verb phrases per clause. However, the ratio thus obtained involves both types of non-finite verbs: infinitives and participles. For the purpose of the current analysis, schemes need to be devised to single out participle clauses for detailed analysis. This can be relatively easily achieved by asking concordancers offered by Wordsmith or AntConc to search for participles that occur in particular syntactic positions within a sentence, aided by manual vetting of the results. In this way, focused

attention can be given to the quantity and quality of student writers' use of participle clauses performing different grammatical functions in comparison with expert use.

Reduced from finite adverbial clauses, participle adverbial clauses mostly have no subject of their own; but when they do, they are called "absolute clauses". The subjectless clauses can be further divided into those that are introduced by a conjunction (Ex. 13) or preposition (Ex. 14) and those that are not (as in Ex. 15 and 16), with the latter referred to as supplementive clauses (Quirk et al., 1985, p. 1123).

(13) ***When asked*** *about her understanding of plagiarism in an interview, Iris presented a standard definition, as shown below…*

(14) ***Upon hearing*** *the beep, the participants were instructed to respond orally in Chinese what they would say in that situation.*

(15) ***Employing*** *the empirical query method they developed, Biber and colleagues identified many MWCs of various lengths…*

(16) ***Influenced*** *by their high-school writing experiences, Lee-Gung, Wei-Shen, and Jun-Yu also chose their arguments for their final research reports at the beginning of the semester.*

When introduced by a conjunction or preposition (as in Ex. 13 and 14) , the participle clause resembles its finite equivalent in terms of its relationship with the matrix clause, but with the tense and aspect of the verb as well as the subject made implicit. Thus, the participle adverbial clause in (13) can be restored as *when she was asked about her understanding of plagiarism in an interview*. Since the production of this type of participle clause is supposedly less challenging due to its explicit semantic relationship with the matrix clause, they are excluded from the current analysis. However, it is interesting to note that Biber et al. (1999, p. 829) found that almost 90% of all adverbial *-ing* clauses and most *-ed* clauses are supplementive clauses, i.e. not introduced by a

subordinator or preposition (as in Ex. 15 and 16). An important usage note for supplementive clauses according to the prescriptive principle is that the reduced underlying subject of the participle clause is supposed to be consistent with that of the matrix clause. When this principle is violated, the writer could face the danger of producing "dangling" or "unattached" participles whose underlying subject differs from that of the matrix clause, although usually implicitly identical with certain elements in it. However, in some circumstances this prescriptive principle can be relaxed (see Biber et al., 1999, p. 892).

The major difference of supplementive clauses from their finite counterparts concerns the degree to which the relationship between the adverbial clause and the matrix clause is explicit. The use of subordinators in finite adverbial clauses automatically makes explicit the relationship but the lack of them in supplementive clauses would more often blur the relationship (Biber et al., 1999), resulting in varying levels of semantic indeterminacy. With no clear semantic category applicable to many instances of supplementive clauses regarding their relationship with the matrix clause, these structures are often taken as "showing a circumstance that supplements the action or state in the matrix clause" (ibid, p. 820).

Similar to finite adverbial clauses, supplementive clauses are also flexible in their syntactic positions, which can appear in the initial, middle, or final position of the sentence. Supplementive clauses used in initial positions, although having indeterminate semantic relationships with the matrix clause, primarily serve to foreground the reason, condition, or time frame for readers to comprehend information contained in the forthcoming matrix clause, for example:

(17) ***Taking a microanalytic perspective****, Lazaraton (2004) studied speech and gestures used by one teacher of English as a second language in her intensive English program.*

When used in middle positions, they appear mostly after the foregoing subject of the matrix clause, performing functions that are "generally indistinguishable from non-restrictive relative clauses" (Kortmann, 1991, p. 9), i.e. to provide additional, relevant information about the subject noun phrase:

(18) *Integrated tests,* ***requiring test takers to use more than one language skill to perform tasks****, may encourage teachers to adopt "a more holistic approach to instruction".*

Final-position supplementive clauses, while remaining semantically implicit as with the initial position, often form a parallel relationship with the matrix clause, resembling a finite coordinate clause introduced by *and*, such as:

(19) *Overall, the VKS form examined demonstrated high reliability under the PCM, although Levels 3 (can give an adequate L1 equivalent) and 4 (can use with semantic appropriateness) appeared to be psychometrically indistinct,* ***operating as alternate methods of testing definitional knowledge***.

Typographically, when used in initial positions, supplementive clauses are preceded by a full stop (or period); when in middle or final positions, they are usually preceded by a comma. Such typographical features of supplementive clauses facilitated corpus-based analysis with concordance tools.

Although absolute participle clauses—those with explicit subjects—were not of a major concern of the present study due to difficulty for corpus search, one type of absolute constructions was included, i.e. participle clauses introduced by *with*, for example:

(20 *All measures showed a high degree of variability and nonlinear patterns of development, **with** **some maintaining roughly the same amount of variation throughout the trajectory**.*

Different from supplementive clauses, absolute clauses permit much greater flexibility in terms of the meaning relationship between the adverbial clause and the matrix clause, in that the explicit subject in the absolute clause is not identical with that of the matrix clause, as shown in Ex. 20. This allows the writer to flexibly provide supplementive but loosely relevant information about certain element in the matrix clause; and the writer does not have to use a coordinate clause or a separate sentence to express the same idea. In addition, the writer is freed from the risk of forming a dangling structure to which supplementive clauses are highly vulnerable, there being no concern about pairing up the subjects of the adverbial and matrix clauses.

As with supplementive clauses, absolute clauses introduced by *with* can also occur in the initial, medial, or final position in the sentence, thus facilitating their identification from concordance search results. Searches can be performed by asking the concordancer to look for instances of *with* appearing after a full stop or a comma, followed by careful manual vetting of the expected clause type.

3.5.3.2. Analytical Procedures for Examining Clause Combination

3.5.3.2.1. Procedure for Examining Finite Adverbial Clauses

Focusing on the semantic categories of reason, concession, and condition, the analysis of finite adverbial clauses first examined their frequency distribution in the two corpora. The results of the distribution analysis can be compared with those of the subordination measures in the L2SCA to see if the two approaches to clause combination yield any consistency in between-corpus differentiation. Rich textual

evidence will be provided for explaining the frequency differences (if any) from the perspectives of discourse organisation and argument development. Where necessary, patterns of students' use of adverbial clauses will be associated with studies on the complexity of argument building in high-quality argumentative writing (e.g. Crammond, 1998), which, among other means, are partly attributed to the use of adverbial clauses.

For each semantic category of the adverbial clauses, only one representative conjunction (or subordinator) was selected for examining its distribution in the two corpora. Biber et al. (1999) found that one single subordinator dominates in each semantic category for academic writing: *because* for causal clauses, *although for* concessive clauses, and *if for* conditional clauses. WordSmith Tools was used to extract all the concordance lines containing these three subordinators, which were then manually vetted to exclude instances where the three words are not used as subordinators of the three semantic categories in question; the raw corpora were used for this analysis.

3.5.3.2.2. Procedure for Examining Participle Adverbial Clauses

The POS-tagged corpora and the concordancer in AntConc were used for identifying the supplementive clauses in different syntactic positions. Sentence-initial instances were preliminarily identified by specifying present and past participles that occur at the beginning of a sentence, i.e. after a full stop; cases where there is an adverb or adverbial phrase in front of the participle were also included by modifying the search syntax.

The results returned by this query would of course include occurrences of present and past participles which do not form supplementive clauses, including 1)

nominal participle clauses and gerunds serving as grammatical subjects, 2) present and past participles as premodifiers of nouns, and 3) prepositions and conjunctions that have been grammaticalised from present and past participles, including but not exclusive to *including*, *excluding*, *concerning*, *regarding*, *considering*, *following*, *barring*, *judging from*/*by*, *according to*, *owing to*, *depending on*, *based on*, *compared with*/*to*, *related to*, *granted* (*that*), *provided/providing* (*that*) and *given* (*that*). The reason for excluding these participles is that using dangling structures is not a problem with these participles-turned-prepositions/conjunctions due to their semantic freedom and thus the meaning relationship between the phrases and clauses of which they are the head and the matrix clause is quite different from the case of participle clauses. Therefore, these irrelevant instances were manually screened for their exclusion from the analysis. To identify sentence-middle and -final supplementive clauses, the search syntax required AntConc to find present and past participles occurring after a comma, with cases having an adverb or adverbial phrase in front of the participle considered as well. Again, irrelevant instances were manually weeded out.

The identification of absolute clauses introduced by *with* was also conducted with the tagged corpora and using the concordancer in AntConc. Firstly, instances of the preposition *with* appearing after a full stop or a comma were automatically extracted from the corpora. Subsequently, the concordance lines were carefully examined to identify eligible cases of absolute clauses. While time-consuming and laborious, this process guarantees accurate and complete inclusion of the expected clause type.

Although these semi-automatic analyses may have left out some eligible cases occurring outside the specified search instructions, the results thus obtained are still highly representative of the expected structures and their distribution across the two corpora.

3.6. Summary of Methodology

This study examined the syntactic complexity of advanced EFL academic writing using a contrastive corpus-based approach, with an aim to identify grammatical and lexical features where student writers significantly lag behind expert writers, which can be taken up as potentially fruitful areas of EAP teaching and learning. The study was divided into two separate, but connected parts: employing an automated syntactic complexity analyser, the first part was dedicated to providing a panoramic view of the various dimensions of syntactic complexity of student writing; the second part examined two specific aspects of syntactic complexity unique to academic writing—noun phrase complexity and the complexity of clause combination.

CHAPTER FOUR

AUTOMATIC MEASUREMENT OF SYNTACTIC COMPLEXITY

4.1. Overview of Chapter

This study is composed of three discrete but related studies of different perspectives of syntactic complexity of EFL academic writing. This chapter provides a general picture of syntactic complexity as reflected by measures derived from second language writing research. The chapter reports descriptive results of 14 syntactic complexity measures computed by the L2 Syntactic Complexity Analyzer (L2SCA) for the two corpora and results of independent samples *t*-tests comparing the means for each of the 14 measures between the two corpora.

Before presenting results of the automatic analysis and comparison of mean scores, the concern about the relationship between subordination and NP-based structures will be re-considered. Since it has been found that development of overall proficiency across academic curricula usually results in a decline in subordination and an increase in clause-internal NP-based expansion, i.e. a trade-off between the two dimensions of syntactic complexity (Byrnes, Maxim, & Norris, 2010; Ortega, 2003), it would be interesting to see if this swap pattern also exists between the two corpora of the present study, representing two distinctly different proficiency populations. If so, there would be a lower representation of NP-based complexity but a higher representation of complexity based on subordination and coordination in the MDC than in the PRAC.

The 14 measures included in the L2 Syntactic Complexity Analyzer are re-presented in Table 4.1.

TABLE 4.1

List of Syntactic Complexity Measures Used in the L2SCA

Measure	Code	Definition
Type 1: Length of production unit		
Mean length of clause	MLC	# of words / # of clauses
Mean length of sentence	MLS	# of words / # of sentences
Mean length of T-unit	MLT	# of words / # of T-units
Type 2*: *Sentence complexity		
Sentence complexity ratio	C/S	# of clauses / # of sentences
Type 3*: *Subordination		
T-unit complexity ratio	C/T	# of clauses / # of T-units
Complex T-unit ratio	CT/T	# of complex T-units / # of T-units
Dependent clause ratio	DC/C	# of dependent clauses / # of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses / # of T-units
Type 4*: *Coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
Sentence coordination ratio	T/S	# of T-units / # of sentences
Type 5: Particular structures		
Complex nominals per clause	CN/C	# of complex nominals / # of clauses
Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units

Note. This table is largely based on Lu (2010, p. 479).

4.2. Results and Discussion

4.2.1. Descriptive Results

Table 4.2 provides the descriptive statistics for the 14 syntactic complexity measures for texts in the two corpora. As can be seen, the mean complexity values of the PRAC outnumber those of the MDC for all the 14 measures except for T/S (T-units per sentence)—a measure of within-sentence clausal coordination, for which the mean value of the MDC is greater than that of the PRAC. Thus, the descriptive means suggest that the MA dissertations appear to be syntactically less complex on all but one of the measures than the research articles. This only reverse case (shaded in grey in Table 4.2), i.e. T/S (# of T-units per sentence), which is a measure of coordinate clausal combination, indicates student writers' slightly greater preference for coordinate clauses than expert writers. With respect to the predicted trade-off between subordination (as measured by DC/C, DC/T) and NP-based structures (partially measured by MLC, CN/T, and CN/C), the descriptive statistics do not show such a pattern across the two corpora, with both structural categories displaying weaker representation in the MDC than in the PRAC. However, the above descriptive comparisons need to be further tested by inferential statistics of means comparison for reliable comments to be made.

TABLE 4.2

Means of Syntactic Complexity Values between the Two Corpora

Measures	Corpus	*N*	Mean	Std. Deviation
Mean length of sentence (MLS)	MDC	70	23.208	3.499
	PRAC	127	26.284	3.904
Mean length of T-unit (MLT)	MDC	70	21.069	2.855
	PRAC	127	24.130	3.411
Mean length of clause (MLC)	MDC	70	12.140	1.370
	PRAC	127	13.192	1.717
Clauses per sentence (C/S)	MDC	70	1.920	.267
	PRAC	127	2.010	.302
Verb phrases per T-unit (VP/T)	MDC	70	2.257	.268
	PRAC	127	2.489	.335
Clauses per T-unit (C/T)	MDC	70	1.741	.188
	PRAC	127	1.840	.224
Dependent clauses per clause (DC/C)	MDC	70	.352	.051
	PRAC	127	.388	.057
Dependent clauses per T-unit (DC/T)	MDC	70	.621	.154
	PRAC	127	.726	.191
T-units per sentence (T/S)	MDC	70	1.100	.055
	PRAC	127	1.090	.060
Complex T-units per T-unit (CT/T)	MDC	70	.431	.079
	PRAC	127	.492	.091
Coordinate phrases per clause (CP/T)	MDC	70	.669	.172
	PRAC	127	.806	.215
Coordinate phrases per T-unit (CP/C)	MDC	70	.388	.107
	PRAC	127	.443	.121
Complex nominals per clause (CN/T)	MDC	70	2.951	.475
	PRAC	127	3.410	.619
Complex nominals per T-unit (CN/C)	MDC	70	1.698	.223
	PRAC	127	1.863	.313

Another consistent pattern discernible from the descriptive statistics is that the standard deviation for each measure in the PRAC is greater than that in the MDC, indicating a wider variation of all the syntactic complexity measures among the research articles and a comparatively more centralised set of values within the dissertations. The greater dispersion of the complexity values in the PRAC suggests that the research articles exhibit a greater variability in expert writers' writing style, with some preferring a more "elaborated" style, e.g. longer clauses or more

subordination, and others favouring the other way around. On the other hand, the smaller within-group variation in the MDC suggests that the values of the syntactic complexity measures do not differ greatly among the dissertations. In other words, student writers are somewhat on a par with each other in their manipulation of the various syntactic devices. Therefore, given the smaller variation of the complexity values in the MDC, the lower mean scores are likely to be representative of all the texts in the corpus. Simply put, the 70 dissertations are relatively equally less complex on all of the 14 measures indicated by the differences of mean scores and standard deviations.

4.2.2. Comparison of Mean Complexity Values

Independent samples *t*-tests were performed to compare the means of the 14 measures for texts in the two corpora to test for statistically significant differences between the two datasets (see Section 3.5.1 for analytical procedure).[30] As can be seen from Table 4.3, the *t*-test results show that a significant difference exists between the two corpora for 12 of the 14 complexity measures ($p < .004$ with Bonferroni correction), indicating that student writing is statistically shown to be syntactically less complex than published writing on these measures. In terms of the five types of syntactic complexity measures conceptualised in the L2SCA (Lu, 2010; see Table 3.3 of this thesis), four out of the five types have shown overall significant difference between the two corpora, with only Type 2 (sentence complexity ratio measured by clauses per sentence) turning out to be not significant. The two measures that do not show any significant difference are both related to clausal coordination, i.e. C/S (clauses per sentence) and T/S (T-units per sentence); thus, the slightly higher use of coordinate clauses in the MDC as shown in the descriptive results turns out not to be statistically

[30] The results shown in this table and their discussions have been published in Liu and Li (2016).

significant. Interestingly enough, these results are highly similar to those found by Ai and Lu (2013), who, employing 10 of the 14 L2SCA measures, also found a significant difference in all but the two clausal coordination measures (see Section 2.6.2 for review of this study).

TABLE 4.3

Independent Samples *t*-Tests of Syntactic Complexity Measures

Measures	T	df	Sig. (2-tailed)	Mean difference	Cohen's *d*
MLS	-5.488	195	.000	-3.076	.786
MLT	-6.376	195	.000	-3.061	.913
MLC	-4.410	195	.000	-1.052	.632
C/S	**-2.077**	**195**	**.039**	**-0.090**	**.297**
VP/T	-4.974	195	.000	-0.232	.712
C/T	-3.155	195	.002	-0.099	.452
DC/C	-4.470	195	.000	-0.037	.640
DC/T	-3.943	195	.000	-0.105	.565
T/S	**1.230**	**195**	**.220**	**0.011**	**.176**
CT/T	-4.699	195	.000	-0.061	.673
CP/T	-4.576	195	.000	-0.137	.655
CP/C	-3.157	195	.002	-0.055	.452
CN/T	-5.393	195	.000	-0.459	.772
CN/C	-3.921	195	.000	-0.166	.562

Note. Significance level was set at $p < .004$ after Bonferroni correction (see 3.5.1 for details of Bonferroni correction).

Some attention to the two clausal coordination measures that do not show a significant difference between the two corpora is needed here: while T/S is a direct measure of coordination ratio, C/S involves the ratio of both dependent and coordinate clauses. Thus, the lack of significant difference on the two measures of clausal coordination indicates that student writers produce coordinate clauses to a similar extent as expert writers. From a developmental perspective (see Section 2.5.2), this finding suggests that clausal coordination does not gain increase as writers' research experience and writing ability develop and therefore it does not deserve much

pedagogical attention. Indeed, since sentences with three or more coordinate clauses would sound awkward and redundant in academic writing, it is necessary for EAP students to be constantly reminded of its excessive use. Indeed, excessive use has been attested in the student corpus, sometimes coupled with the presence of infelicitous use of run-on coordination, such as:

> *Teachers can well transfer their feelings to students during teaching and learning, and students will in turn feel that they are respected and emphasized, thus the class atmosphere can be promoted to be free and at ease.* (Text 14, MDC)
>
> *By examining a great number of original student texts, we may find out typical grammar errors made by Chinese students in their writing, then categorize them, and discover the most frequent types with their proportions and characteristics.* (Text 24, MDC)

However, it should be noted that while the measures of clausal coordination have not shown significant differences, both the two measures of phrasal coordination, i.e. CP/C (coordinate phrases per clause) and CP/T (coordinate phrases per T-unit), have attested a significant difference, indicating that student writers would tend to produce fewer coordinate phrases than expert writers. This finding makes sense considering that "academic writing contrasts strongly with conversation in favouring phrase-level coordination, which helps to build up complex embedded structures" (Biber, Conrad, & Leech, 2002, p. 228) and that in academic writing the word *and*, the most frequent coordinator, is more typically used as a phrase-level connector rather than as a clause-level connector (Biber et al., 1999). Thus, although the two corpora appear to be "equally complex" on measures of clausal coordination which is more typical of spoken registers, student texts are less complex than published texts on

measures of phrasal coordination which is more characteristic of academic writing. In addition, this finding also suggests that the two measures involving clausal coordination (T/S and C/S) cannot adequately differentiate between student writing and expert writing representing two distinct academic literacy levels. This is consistent with the results reported in most previous studies where T/S has not been shown to discriminate curriculum levels (cf. Lu, 2011). Thus, caution should be exercised in selecting clausal coordination measures for studies aiming to track the development of students' syntactic complexity in academic writing across curriculum levels.

4.2.3. Relationship between Subordination and Noun Phrase Complexity

The answer to the question whether a predictable trade-off pattern exists between clause-internal noun phrase complexity and subordination complexity appears to be negative in the present study (see Section 2.5.2). Although there is a significantly higher representation of NP-related complexity (as measured by MLC, CN/C, and CN/T) in the expert corpus, expert writing also shows a significantly greater use of subordination (as measured by C/T, DC/C, and DC/T). Thus, the anticipated trade-off effect among these two types of syntactic complexity has not been detected in this study. A plausible reason for this is that the "curvilinear relationship" (Ortega, 2003, p. 514) between amount of subordination and NP expansion found in previous studies exists across curriculum levels (e.g. Lu, 2011) and that the proficiency gap between the two groups of writers in this study is so large that the curvilinear relationship has reached a point where the two variables in the relationship (i.e. subordination and NP expansion) might both begin to increase again.

As such, the developmental prediction (Ortega, 2003; Sharma, 1980; Wolfe-Quintero et al., 1998) arguing for more NP-based and less subordination-based

complexification along with advancement to higher proficiency levels is not adequately supported in this study which compares MA dissertations and published research articles. Similarly, the reason why Ai and Lu's (2013) study comparing the writing of college-level EFL students and native speakers has also failed to identify such a relationship may also be due to the excessively wide proficiency gap between the two writer groups in their study.

But whatever statistical fluctuations there could be between subordination and NP-related structures, it has been clearly shown that the use of dependent structures in academic writing indeed constitutes an important area for advanced student writers. Although Biber et al. (2011) have found that finite dependent clauses are used to a much lesser extent in academic writing than in conversation (except for post-nominal relative clauses), the roles they play in establishing the relationships among propositions and in allocating information status within the discourse cannot be overlooked while emphasising noun phrase complexity as the most conspicuous (or frequent) feature of academic discourse. These two dimensions of syntactic complexity are not necessarily mutually exclusive of each other: whereas NP-based structures contribute to grammatical complexity by lengthening the clause and in turn the whole sentence, many types of dependent clauses add to complexification more the logico-semantic level (Halliday & Matthiessen, 2004). From the perspective of text comprehension, length-induced complexity places the burden on the reader's working memory capacity and subordination-induced complexity would call for readers' cognitive ability and domain-specific knowledge to deduce inter-clausal meaning relations. Of course, an interaction of these two types of complexity would be cognitively taxing and would just heighten the overall challenge for text comprehension.

4.2.4. Justifying Measures of Subordination

The finding of significantly greater use of subordination in expert writing has in some way refuted criticisms levelled against using subordination-based measures to study syntactic complexity in academic writing (see Section 2.4.2). At the same time, the pedagogical legitimacy of further enhancing students' development in this respect in EAP writing instruction has been justified. However, the challenge for identifying and producing dependent relations in writing is not just a purely grammatical issue—it also requires the writer's keen familiarity with and insightful understanding of the specific research domain in question. For example, an important role of using finite adverbial clauses is to build argumentation and provide justification, which can only be based on the writer's involved engagement with the area being written about (Li, Hyland, & Hu, 2017). Previous research has found that content familiarity has a positive effect on syntactic complexity and critical thinking in L2 writing which are usually measured by the use of circumstantial adverbial clauses encoding such logical relations as concession, reason, and contrast (Stapleton, 2001; Tedick, 1988).

However, measures of overall subordination as those included in the L2SCA are inadequate for a comprehensive picture of how functionally different subordinate clause types contribute to the final results. As discussed in 2.5.1, three types of finite dependent clauses that serve as important discourse unfolding devices in academic writing are subsumed in the measures of subordination: embedded complement (or nominal) clauses and restrictive relative clauses on the one hand and hypotactic adverbial clauses and non-restrictive relative clauses on the other. Therefore, all three clause types need to be taken account of when interpreting the finding that expert writers use more subordination than student writers. Unfortunately, however, it is impossible to deduce this knowledge from the results of the overall subordination

measures. Complement *that*-clauses, for example, are frequently taken by a wide range of reporting verbs and predicative adjectives in academic writing, realising a number of discourse functions (e.g. Hyland & Tse, 2005). This structure is especially useful in conveying the author's stance towards and evaluation of the content expressed in the *that*-clause. Apart from those introduced by *that*, complement clauses introduced by *wh*-complementisers (e.g. *what*, *how*, and *why*) are also useful and versatile in writing, being able to fill in most syntactic slots that allow nominals. On the other hand, among the common semantic categories of finite adverbial clauses (time, manner, reason, concession, and condition) provided by Biber et al. (1999), clauses illustrating reason, concession, and condition are important devices to present justification and argumentation, which are central to the objectives of academic writing. Thus, measures of subordinate clauses that better differentiate the above discussed clause types are needed. Even more fine-grained analyses of differential semantics within the same clause type are necessary in order to understand which specific meaning-making categories are insufficiently used by student writers.

The above brief recount of the functions of dependent clauses in English academic writing is intended to highlight them as an important avenue of EAP students' ongoing syntactic complexity development along with developing clause-internal, noun phrase complexity. Students need to be made aware that NP complexity requires integrating information canonically conveyed by clauses via such devices as nominalisation and grammatical metaphor. However, it would also be important for them to realise that information packing cannot go unlimited and has to stop at a certain point so that meaning is effectively conveyed and understood. This is where dependent structures come in, bringing a logical dimension to the heavily packed clauses, such as the following sentence taken from the PRAC:

Although explicit knowledge of grammatical rules ***does not guarantee*** *corresponding changes to implicit grammatical accuracy, research since the mid1980s* ***has suggested*** *an indirect, but facilitative, role for explicit knowledge in improving the rate and ultimate level of L2 attainment*.

From this example we can see that both the dependent clause and the main clause are characterised by highly condensed NP structures (underlined) connected by lexical verbs (in bold) for their meaning relationships. But there is also a kind of complexity at the inter-clausal level, that is, the concessive orientation of the dependent clause headed by *although* to the main clause containing the author's intended argument. It is this concessive relationship that is challenging for the writer to identify and the reader to construe the relationship between the two propositions, given the already highly packed NP structures. Thus, the sentence is further complexified by the confluence of the cognitively demanding concessive relationship conveyed through subordination.

Another limitation of what can be interpreted from the results of the subordination-based measures is that they can only capture finite dependent clauses (Lu, 2010), leaving non-finite dependent clauses (e.g. infinitive and participle clauses) subsumed under the verb phrase measure—VP/T (verb phrases per T-unit) and the two measures of complex nominals—CN/C (complex nominals per clause) and CN/T (complex nominals per T-unit).[31] With VP/T tapping into both finite and non-finite verbs and with CN/C and CN/T involving infinitival and participle structures, these measures cannot provide accurate information on the use of non-finite clauses in the text. This problem has been recognised by Yang et al. (2015), who made some adaptations to the L2SCA for directly dealing with non-finite structures in their study.

[31] Gerunds and infinitives in subject position are counted as "complex nominals" in the L2SCA (Lu, 2010, p. 483).

Biber et al. (1999) shows that academic writing uses non-finite clauses much more frequently than conversation and Biber et al. (2011) have further revealed that all subtypes of non-finite clauses except two post-verbal types (i.e. verb + infinitive and verb + -ing) are significantly more frequently used in academic writing than in conversation. Therefore, it is suggested that research into students' syntactic complexity development should examine the extent to which non-finite clause types favoured in academic writing are used by EFL students, especially participle clauses functioning as subjects, adverbials, and postmodifiers (Biber et al., 1999; Granger, 1997).

4.2.5. Focusing on Noun Phrase Complexity Measures

The three measures tapping into NP-based complexity—MLC, CN/C, and CN/T—have all displayed significant differences between the two corpora, indicating a more elaborated nominal complexity in expert writing than in student writing. However, the two measures of complex nominals, by definition, involve three structural types of a more-or-less nominal nature: 1) noun phrases with pre- or post-modifiers, 2) nominal clauses, and 3) gerunds and infinitives in subject position. Only the first type is concerned with a strictly defined noun phrase complexity, with the other two falling within the scope of clausal structures. Therefore, these measures cannot provide straightforward information on the extent to which NP-based complexity is demonstrated across the two corpora.

Motivated by the need to specifically examine noun phrase complexity derived from multiple postmodification, Yang et al. (2015) modified the original pattern used to identify complex nominals to devise an independent measure of the ratio of complex

noun phrases.[32] Knowing the density of complex noun phrases can serve as a point of departure for more in-depth investigations into noun phrase complexity. However, such tasks would best be performed by a human analyser. A number of dimensions to this more qualitative venue of investigation can be examined, including length and depth of postmodification, structural types of postmodifiers, discourse functions of postmodification, and semantic ambiguity and verbosity in the postmodifier. These features will be expressly taken up in Chapter 4.

4.3. Summary of Chapter

This chapter has examined different dimensions of the syntactic complexity of student writing as compared with expert writing, on 14 measures included in the L2 Syntactic Complexity Analyzer. It aims to see which of these syntactic complexity measures between the two groups of writers reach statistical significance, with a secondary goal to see how nominal structures and subordinate constructions are represented in the two groups. It has been found: 1) syntactic complexity of student writing is significantly weaker than that of expert writing attested on all but two measures—T/S and C/S, both capturing something of clausal coordination within the sentence; and 2) there does not appear to be a trade-off, or a curvilinear relationship between the use of subordinate structures and nominal structures as hypothesised by the developmental prediction (Ortega, 2003; Wolfe-Quintero et al., 1998), suggesting the importance of both noun phrase complexity and subordinate-based complexity for the development of advanced EFL academic writing.

[32] This specific definition of complex noun phrase (CNP) does not include postmodifiers realised by relative clauses—both finite and non-finite, to avoid overlaps with other measures used in that study (Lu, personal communication).

However, the results returned by the L2SCA measures need careful interpretation. This is because definitions of the syntactic units of the L2SCA measures have followed the tradition of composition research, and they are sometimes found to be at odds with the definitions in grammatical analysis (Yang et al., 2015). For example, dependent clauses as defined in the subordination-based complexity measures only include finite clauses, disregarding non-finite dependent structures which are usually categorised as dependent clauses by most grammatical frameworks. Thus, non-finite dependent clauses are excluded from subordination-based complexity but subsumed under verb phrase measures. In addition, the three components subsumed under the complex nominals measures and subordination measures also need to be examined in their own right to provide different perspectives on the use of nominal structures, particularly complex noun phrases with multiple postmodification.

In light of the limitations of the L2SCA measures to unravel how particular dimensions of syntactic complexity truly work, this study performs a detailed examination of two of the major dimensions of syntactic complexity: noun phrase complexity and clausal complexity. In the next chapter, different aspects of noun phrase complexity in student writing—which is actually one of the three components of the complex nominals measures—will be examined. Findings from these analyses are expected to provide complementary and interpretive support for the syntactic complexity measures and therefore to better inform EAP instruction.

CHAPTER FIVE

NOUN PHRASE COMPLEXITY IN STUDENT WRITING

5.1. Overview

This chapter reports and discusses different aspects of noun phrase complexity in student writing as one of the most distinctive grammatical features of academic discourse. This is motivated by the inability of complex nominals measures to unravel how different types of nominal structures work, the use of complex noun phrases in particular. Therefore, examining features contributing to the expansion of the noun phrase, especially those that complexify NP postmodification, need to play a central role in such an endeavour.

As described in 3.5.2, the examination of noun phrase complexity included structural types of postmodifiers, length and depth of multiple postmodification, and the use of complex noun phrases as clause subjects. Quantitative results are further contextualised in discourse-informed analyses for demonstrating the significance of NP complexity to the establishment of discourse coherence.

The examination of noun phrase complexity is based on NP samples expanded from a set of head nouns. The analyses reported in 5.2.1 and 5.2.2 were performed on NPs headed by *development*, *information*, *relationship*, *difference*, *analysis*, and *knowledge*, while those in 5.2.3 through 5.2.5 on NPs headed by *understanding*, *analysis*, and *lack* with the first postmodifier being an *of*-PP.

5.2. Results and Discussions

5.2.1. Nouns with Postmodification in General

This section reports the tendency of the six head nouns to take postmodification in the two corpora in proportional terms. Attention will be paid to the extent to which student writers tend to postmodify nouns differently from expert writers and the tendency of different head nouns to take postmodification (see Section 3.5.2.3 for analytical procedure). The proportion of NPs with postmodification for the set of NP samples as a whole is presented first, followed by the results for NPs with each individual head noun.

There are altogether 6,192 instances of the six head nouns in the PRAC and 7,776 in the MDC, of which 2,573 (42%) are postmodified in the former and 3,257 (42%) in the latter. Thus, it can be seen that student writers are on a par with expert writers in the tendency to use postmodification to expand abstract nouns. While it might be a coincidence that the proportions of postmodified nouns by the two groups of writers are strikingly similar, the results show that postgraduate student writers have achieved expert-like consciousness in providing relational, delimiting information about nominalised processes, concepts, or properties. Student writing is sometimes found to feature an even denser use of postmodification than expert writing, as the following two extended extracts from the two corpora show (with nouns having at least one postmodifier marked in bold):

> (1) *The **distinctiveness** of writing **item** of CET 4, the national foreign language **examinations** for professional advancement, is also worth discussing. Firstly, it is the **writing** for examination, which means the **readers** of the students' writings are **raters** who are, in most cases, advanced writers. Secondly, it is designed with controlled **length** on essay (about 120 words) and on time*

allocation (Students are supposed to finish their composition within 30 minutes). Thirdly, different ***types*** *of writing-task do pose different* ***levels*** *of difficulties for learners. The task* ***types*** *of CET 4 writing are mainly discursive (explanatory/argumentative) and these tasks involve high* ***levels*** *of idiomatic, phrasal verb* ***knowledge*** *plus the* ***ability*** *to generate complex sentences. The above mentioned factors, with the* ***complexity*** *of writing itself, make it more difficult for* ***students*** *who could be regarded as lower-intermediate* ***learners*** *from vocational institutes. Hereby, it makes sense that students take the shortcut by copying the model sentences.* [152 words] (Text 61, MDC)

(2) *In addition, owing to the* ***exodus*** *of* ***emigrants*** *into English-speaking countries (the United States in particular), since the early 1980s, there has been a subsequent* ***return*** *of overseas* ***Chinese*** *whose Chinese language has been noticeably influenced by the English language not only in vocabulary and grammar but in pragmatics. They have retained* ***remnants*** *of Anglo-American influence not only when conversing with each other, but also in their* ***communication*** *with local residents, since this supports their* ***reputation*** *as foreign-language-speaking cosmopolitans (for the most part accompanied by certain wealth). Therefore, the* ***influence*** *of English on Chinese has progressed further in extent and increased in intensity at the current* ***stage*** *of language development. Not only has it accumulated a large body of* ***information*** *on lexical borrowing, morphological transference and syntactic variation, but as we have seen in this study, also influenced the* ***pragmatics*** *of the Chinese language. A good* ***case*** *in point is the increasing* ***use*** *of appreciation CR strategy in Chinese speakers. Yuan (2002) also offers* ***discussions*** *about the* ***change*** *of*

culture in China, arguing that the social changes are the ***reasons*** *for her subjects to adopt more acceptance, but less non-acceptance strategies than previous* ***studies*** *on Chinese CRs.* [197 words] (Text 16, PRAC)

With 16 nouns out of 152 words having a postmodifier in Ex. 1 (i.e. 10.5%) and 17 nouns out of 197 words having a postmodifier in Ex. 2 (i.e. 8.6%), Ex. 1 is denser than Ex. 2 in terms of postmodification frequency. With other features left unaccounted for, examining the extent to which nouns are postmodified alone does not reveal much material difference between student writing and expert writing. These other features include the structural types, length, and depth of postmodification, the results of which will be presented and discussed in subsequent sections of this chapter.

On another note, although highly comparable in the overall extent of postmodification between the two corpora, when the six nouns are examined individually, there are variations in the extent to which nouns are postmodified, as shown in Figure 5.1.

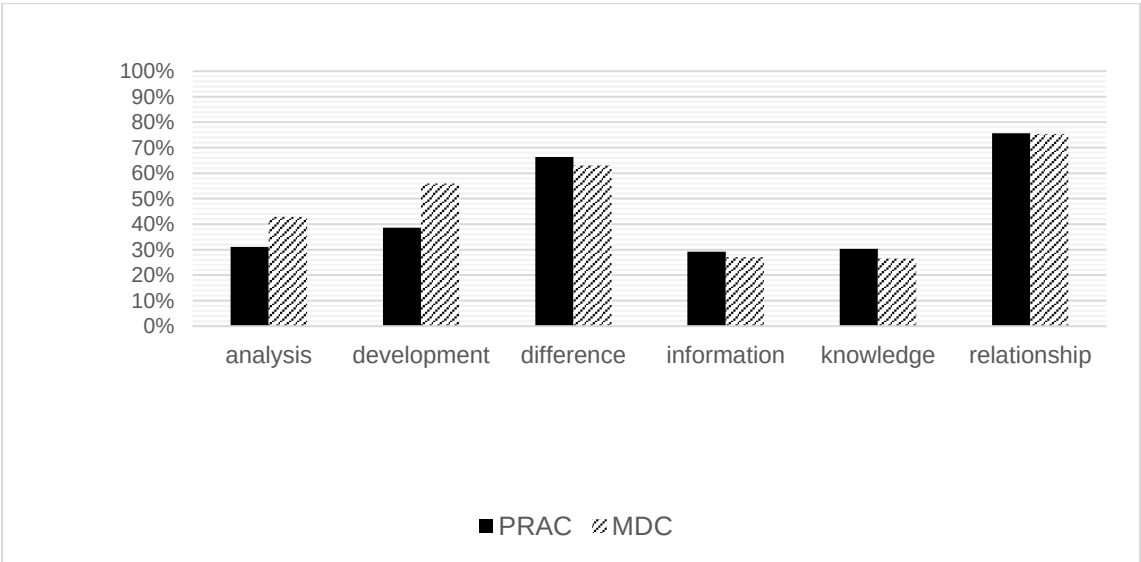


Figure 5.1. Proportions of NPs with postmodification for each noun.

As can be seen, some nouns (*analysis* and *development*) are more frequently postmodified in the MDC while some (*difference, information,* and *knowledge*) in the PRAC. It can also be seen that the gaps between the two corpora in using postmodification for *analysis* and *development* are larger than those for *difference, information,* and *knowledge*; there is almost no difference for *relationship*. Another important finding is that for both corpora some nouns are more prone to postmodification than others, with the proportions ranging from as low as 27% (*knowledge*) to as high as 76% (*relationship*), quite a distance from the average of around 42%. This shows that nouns differ considerably in such a colligational feature, probably due to their differing semantic properties and hence the necessity for them to be postmodified and the variety of postmodifying structural types.

It is interesting to see that student writers perform more or less the same as expert writers in the extent to which each noun takes postmodification, despite there being an obvious difference in the postmodification for some nouns. In this sense, the almost equivalent overall proportions of postmodification in the two corpora can be explained by the roughly similar individual instantiations. However, it may not seem that surprising in light of arguments based on frequency effects on language acquisitions that "[f]requency of occurrence may lead to independent representation of even so-called regular constructional patterns" (Ellis, 2002, p. 168)—if postmodification can be regarded as a kind of "regular constructional patterns". This effect may be stronger for high-frequency nouns such as those selected for analysis in the present study, hence their similar postmodification proportions between the two corpora. However, the extent to which students also show expert-like performance on less frequent nouns is not known from the present analysis and needs to be further researched.

Despite the striking similarity in the overall proportions of postmodified nouns between the two corpora, it would be more important to know the extent to which student writers use different structural types of postmodifiers in comparison with expert writers. The following sections present the results of these comparisons in proportional terms as described in the analytical procedure.

5.2.2. Postmodifiers of Different Structural Types

This section presents the results of various postmodifying structural types in the two corpora in proportional terms. The analysis examined three pairs of contrasting structural types that are conceptually mutually exclusive (i.e. PPs vs. clauses, PPs headed by *of* vs. "other" PPs, and finite clauses vs. non-finite clauses) by examining the proportion of one structural type in each dichotomous pair among occurrences of the superordinate structural type, e.g. the proportion of non-finite clauses among occurrences of all clausal postmodifiers. The process was then repeated for examining NPs with each individual head noun (see Section 3.5.2.3 for analytical detail). First reported are the results for NPs headed by the six nouns as a whole and then those for each individual head noun are presented. As set out in 3.5.2, three pairs of conceptually contrasting structural types are compared between the two corpora: phrasal vs. clausal postmodification, PPs headed by *of* vs. other PPs, and finite clauses vs. non-finite clauses.

5.2.2.1. Phrasal vs. Clausal Postmodification

Figures 5.2-5.4 present the relative proportions of phrasal versus clausal postmodifiers in the two corpora. As shown in Figure 5.2, both student writers and expert writers use preposition phrases as the absolute major type of postmodifier,

accounting for around 93% in the expert corpus and 90% in the student corpus. Postmodification by clauses, in contrast, only take on a marginal representation, 7% in the PRAC and 10% in the MDC. While the two corpora are generally similar on these comparisons, the student corpus has shown a slightly higher use of clausal postmodification and lower use of prepositional phrases. This appears to be consistent with Biber et al's (2011) hypothesis that phrasal postmodification develops at a later stage than its clausal counterpart (Stage 3-5 vs. Stage 3 & 4). However, the gap between the two groups of writers in this study is quite small, indicating that student writers have almost attained expert-like performance in choosing between clausal and prepositional postmodification.

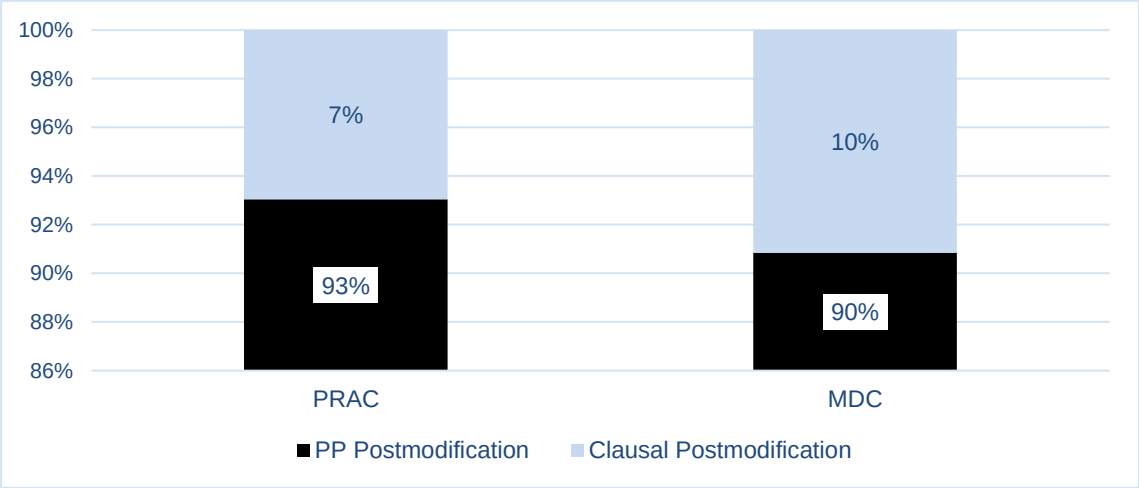


Figure 5.2. Proportions of postmodification by prepositional phrases versus clauses.

Just as nouns differ in the extent to which they tend to take postmodification, they are also somewhat different in their preferences for clausal and phrasal postmodifiers and this difference is evident in both corpora (see Figure 5.3 and 5.4). Although being the predominant type of postmodification for all six nouns, preposition

phrases are more closely associated with some nouns (e.g. *relationship*, *difference*, *development*) than with others (e.g. *information* and *knowledge*). It can be seen that among the six nouns, *information* is least modified by preposition phrases, followed by *knowledge*. This noun also sees the biggest gap between the two corpora in the extent of taking prepositional and clausal postmodifiers; all the other nouns only show little or no difference. With almost 40% of the postmodifiers of *information* being clausal structures in the MDC (as opposed to only 22% in the PRAC), students seem to be more primed to clausal structures to go with this noun. Thus, the overall 93% vs. 90% difference (Figure 5.2) may mainly derive from the big gap with *information*. It can be therefore inferred that postgraduate students have attained expert-like awareness in using phrasal and clausal postmodification but they sometimes seem to prefer clausal structures for certain nouns, indicating advanced L2 writers' still growing productive phraseology with respect to prepositional postmodification.

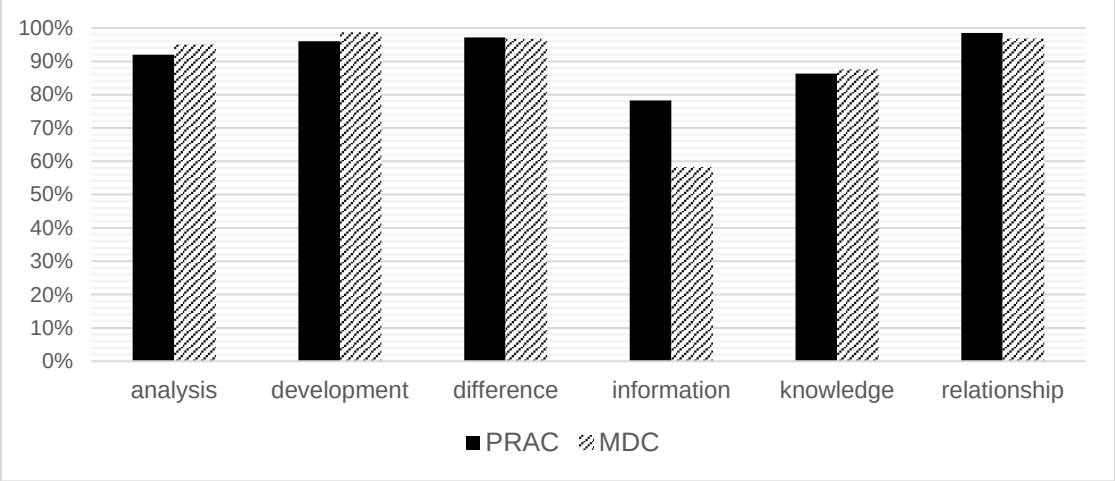


Figure 5.3. Proportions of postmodification by preposition phrases for each noun.

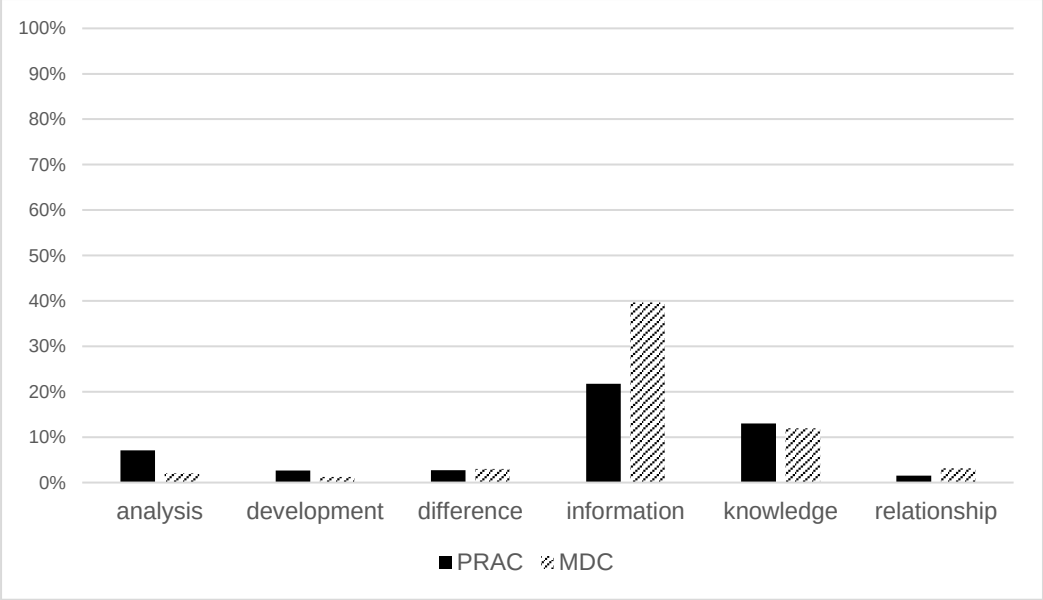


Figure 5.4. Proportions of clausal postmodification for each noun.

5.2.2.2. Postmodification by Two Types of Preposition Phrases

Figures 5.5–5.7 present the use of two types of prepositional postmodifiers for the NP samples as a whole and for individual head nouns in the two corpora: PPs headed by *of* and those headed by prepositions other than *of*. As can be seen from Figure 5.5, student writers use more *of*-PPs and fewer "other" PPs than expert writers, with both types of PP postmodifiers evenly represented in student writing. This general difference may indicate that student writers use a smaller variety of prepositions besides *of* in PP postmodification. Since prepositions other than *of* encode a vast array of concrete/locative and abstract meanings, using more of such PPs is conducive to creating a broader spectrum of physical and logical relationships between the postmodifier and the head noun. But it would be too hasty to conclude that student writers lack the ability to encode such a spectrum just based on the small NP samples of this study because some of the six nouns take as their canonical postmodifiers PPs with prepositions other than *of*, such as *difference* and *relationship*, both predominantly taking *between*-PPs. A detailed examination of the prepositional postmodifiers for each

individual head noun would be more revealing about students' performance in this respect.

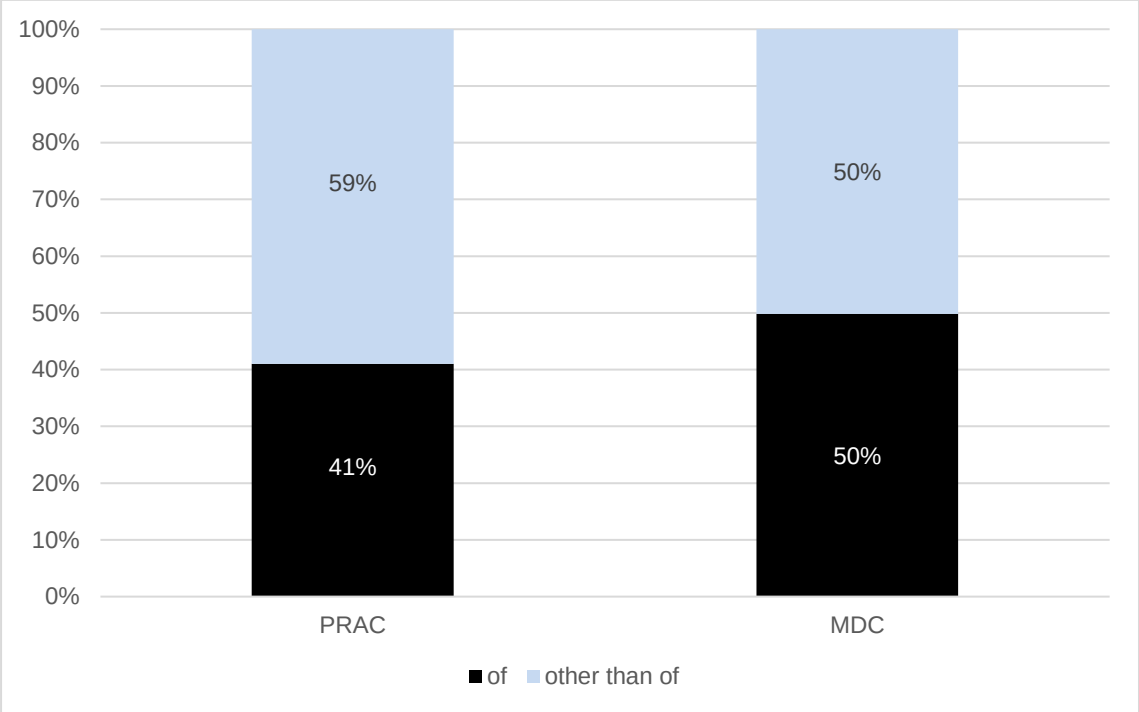


Figure 5.5. Proportions of two types of prepositional postmodification.

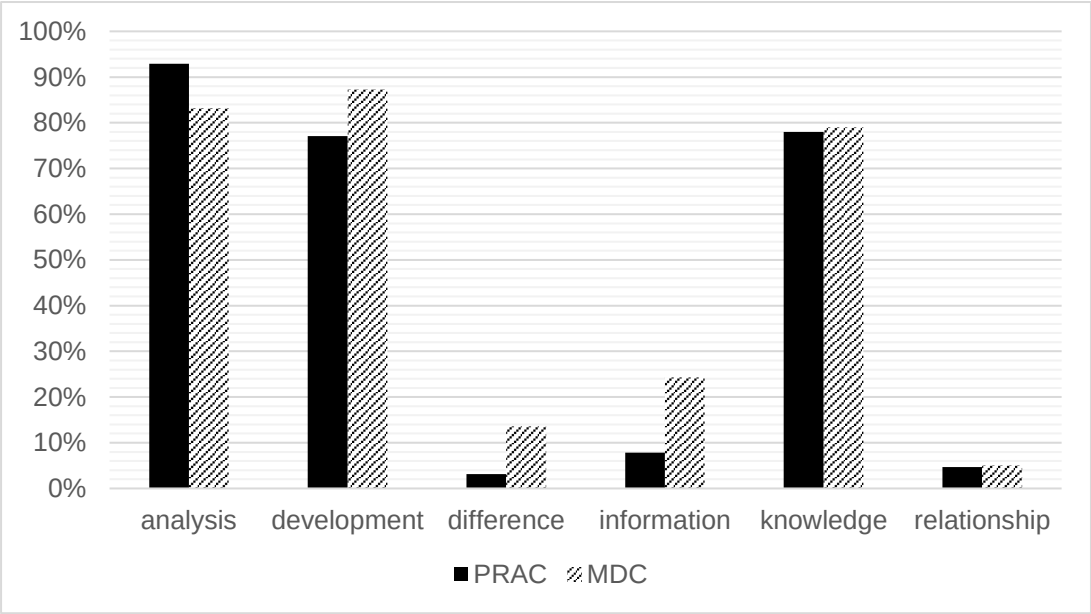


Figure 5.6. Proportions of PP postmodifiers headed by *of*-phrases for each noun.

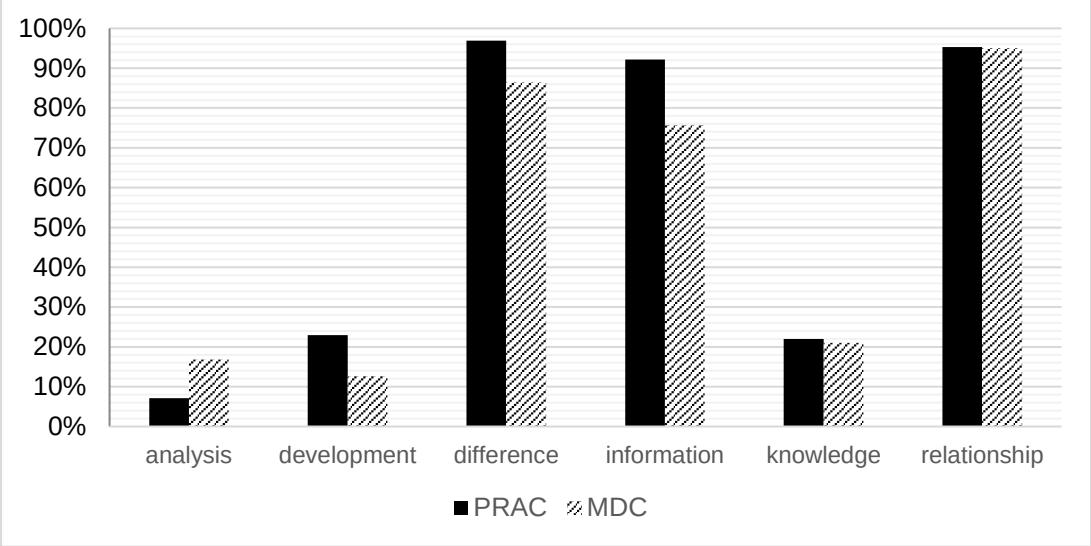
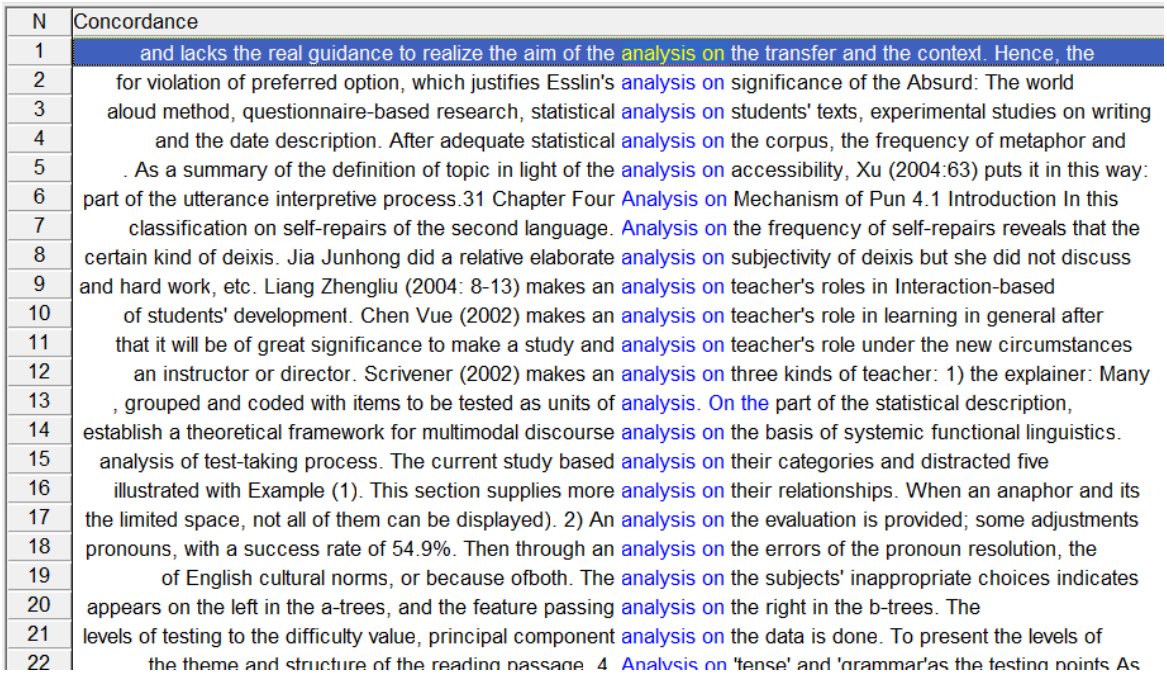


Figure 5.7. Proportions of PP postmodifiers headed by prepositions other than *of* for each noun.

Scrutinising the types of postmodifying PPs for each individual head noun, it can be seen that student writing conforms to expert writing in terms of the predominant PP type associated with each noun. Specifically, *analysis*, *development*, and *knowledge* predominantly take *of*-PPs while *difference*, *information*, and *relationship* take PPs with other prepositions. However, the two corpora differ in the exact extent of the six nouns taking the two PP types, with only *knowledge* and *relationship* showing equivalent performances where *knowledge* is predominantly followed by *of*-PPs and *relationship* by "other" PPs. The results seem to show that student writers tend to overuse *of*-PPs and underuse "other" PPs for *development*, *difference* and *information*. This pattern may be able to explain its slightly lower representation in the whole NP sample from the student corpus (Figure 5.5.).



N	Concordance
1	and lacks the real guidance to realize the aim of the analysis on the transfer and the context. Hence, the
2	for violation of preferred option, which justifies Esslin's analysis on significance of the Absurd: The world
3	aloud method, questionnaire-based research, statistical analysis on students' texts, experimental studies on writing
4	and the date description. After adequate statistical analysis on the corpus, the frequency of metaphor and
5	. As a summary of the definition of topic in light of the analysis on accessibility, Xu (2004:63) puts it in this way:
6	part of the utterance interpretive process.31 Chapter Four Analysis on Mechanism of Pun 4.1 Introduction In this
7	classification on self-repairs of the second language. Analysis on the frequency of self-repairs reveals that the
8	certain kind of deixis. Jia Junhong did a relative elaborate analysis on subjectivity of deixis but she did not discuss
9	and hard work, etc. Liang Zhengliu (2004: 8-13) makes an analysis on teacher's roles in Interaction-based
10	of students' development. Chen Vue (2002) makes an analysis on teacher's role in learning in general after
11	that it will be of great significance to make a study and analysis on teacher's role under the new circumstances
12	an instructor or director. Scrivener (2002) makes an analysis on three kinds of teacher: 1) the explainer: Many
13	, grouped and coded with items to be tested as units of analysis. On the part of the statistical description,
14	establish a theoretical framework for multimodal discourse analysis on the basis of systemic functional linguistics.
15	analysis of test-taking process. The current study based analysis on their categories and distracted five
16	illustrated with Example (1). This section supplies more analysis on their relationships. When an anaphor and its
17	the limited space, not all of them can be displayed). 2) An analysis on the evaluation is provided; some adjustments
18	pronouns, with a success rate of 54.9%. Then through an analysis on the errors of the pronoun resolution, the
19	of English cultural norms, or because ofboth. The analysis on the subjects' inappropriate choices indicates
20	appears on the left in the a-trees, and the feature passing analysis on the right in the b-trees. The
21	levels of testing to the difficulty value, principal component analysis on the data is done. To present the levels of
22	the theme and structure of the reading passage. 4. Analysis on 'tense' and 'grammar'as the testing points As

Figure 5.8. Snapshot of concordance lines for analysis on in the MDC.

Observing students' actual use of PP postmodifiers, it has been found that sometimes students' erroneous selection of prepositions may lead to overuse and underuse of a certain type, not least when the misuses occur across the whole corpus. To illustrate, 53 cases of "*analysis on*" have been found in the student corpus as opposed to only two in the expert corpus. Figure 5.8 is a snapshot of the concordance lines of this bi-gram in the student corpus. While the use of *on* to go with *analysis* cannot be said to be absolutely wrong, it does sound slightly unidiomatic and is too infrequent in the expert corpus to be seen as standard use. As can be seen from the concordance lines, many instances of this bi-gram would mean the same as *analysis of*. This could be caused by negative transfer from the equivalent Chinese expression "关于 XX 的研究", where "关于" is usually rendered through the preposition *on* in translation practices. Another case is students' use of the bi-gram *difference of*, which occurs 101 times in the MDC but only 22 times in the PRAC. Examining the

concordance lines, it is found that many instances of this bi-gram in the student corpus are actually specious uses of *difference in*, as shown in the following example:

(3) *What are the similarities and* ***differences of the use of "you know"*** *between native speakers and Chinese English majors…*(Text 05, MDC)

However, *difference of* in the expert corpus is more often limited to the description of a quantitative gap where the *of*-PP serves as a complement substantiating the shell noun *difference*, as in the following:

(4) *For Bohai TS the proportions are 69.67% and 30.33****%, a difference of 39.34 percentage points***. (Text 66, PRAC)

Again, using *of* instead of *in* to go with *difference* is not necessarily wrong given its high degree of semantic versatility and opaqueness. However, efforts need to be made for clear meaning presentation when there are better resources available, in this case opting for *in* with *difference*. Finally, the exhibited overuse of *information of* is attributable to the same problem causing the previous two overused bi-grams. Similarly, concordance lines of *information of*, which occurs 61 times in the MDC, show that many of its instances are where expert writers would have used *information on*/*about*.

The analysis of inaccurate selections of prepositions has in part helped explain the overuse or underuse of particular types of PP postmodifier. This has also demonstrated the kind of cognitive difficulty associated with the selection of prepositions as decided by the semantics between the two nouns linked by the preposition. It is also worth noting that lexicogrammatical patterns in student writing are often subject to error-induced influences, so their comparisons with expert texts should be interpreted with this concern taken account of.

5.2.2.3. Finite Relative Clauses vs. Non-finite Relative Clauses as Postmodifiers

Figures 5.9-5.11 diagrammatically present the use of finite and non-finite relative clauses in the two corpora for the NP samples as a whole and for each individual head nouns. It can be seen from Figure 5.9 that student writers use a higher amount of finite relative clauses and therefore a smaller proportion of non-finite clauses than expert writers. This finding is consistent with previous research that reveals L2 writers' underdeveloped use of non-finite clauses, both adverbial and relative alike (e.g. Granger, 1997). It also lends support to Biber et al.'s hypothesis that non-finite relative clauses develop at a later stage than their finite counterparts. However, the six head nouns have shown both consistent and contrary patterns regarding the general difference. Specifically, consistent with the overall pattern, student writers use more finite relative clauses to postmodify *development*, *difference*, and *knowledge* but substantially more non-finite clauses for *relationship*; there is little to no difference with *analysis* and *information*. Thus, the overall between-corpus difference (Figure 5.9) should be seen as a result of differences with NPs of certain head nouns, yet incapable of accounting for all head nouns.

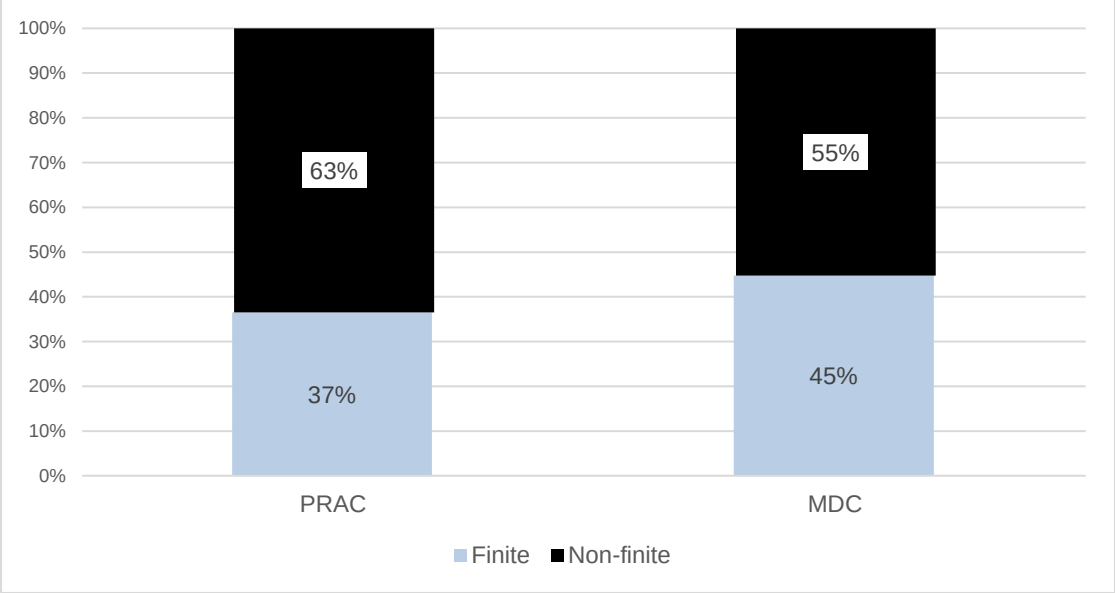


Figure 5.9. Proportions of clausal postmodification by finite clauses versus non-finite clauses.

Especially atypical of the overall trend is the use of finite/non-finite clauses for NPs headed by *relationship* by the two groups of writers, the pattern of which marks a stark contrast with NPs headed by the other head nouns. Student writers, for some reason, have a totally different appreciation of the type of clausal postmodifiers to co-occur with this noun. However, it needs pointing out that both groups of writers use relatively few clausal postmodifiers for all six head nouns, with six in the PRAC and 13 in the MDC postmodifying *relationship*. With so little clausal postmodification used for this head noun, the claim that student writers have a strong preference for finite postmodifiers needs to be somewhat moderated. It is also small wonder that the overall trend (Figure 5.9) would not be much affected by the large gap found in *relationship*.

A possible reason for the student corpus to show greater finite clausal postmodification may stem from student writers' fondness for *which*-clauses, particularly for *information* and *knowledge*. It is found that 17 *which*-clauses are used to co-occur with *information* in the MDC versus none in the PRAC and 14 are used for

knowledge in the MDC versus 3 in the PRAC. It is not clear what has contributed to students' preference for relative clauses introduced by *which* vis-a-vis *that* and zero relativiser. Apart from restrictive relative clauses being examined in this chapter, student writers have also been found to overuse non-restrictive *which*-clauses as a device of clause combination (see 6.2.3). Although previous research has found Chinese EFL students (estimated to be at upper-intermediate proficiency level) to underuse relative clauses as a result of avoidance strategy induced by fear of error (Schachter, 1974), the findings of this study call on future research to explore pedagogically and/or psycholinguistically related reasons for advanced Chinese students' prolific use of non-finite relative clauses in general and *which*-clauses in particular.

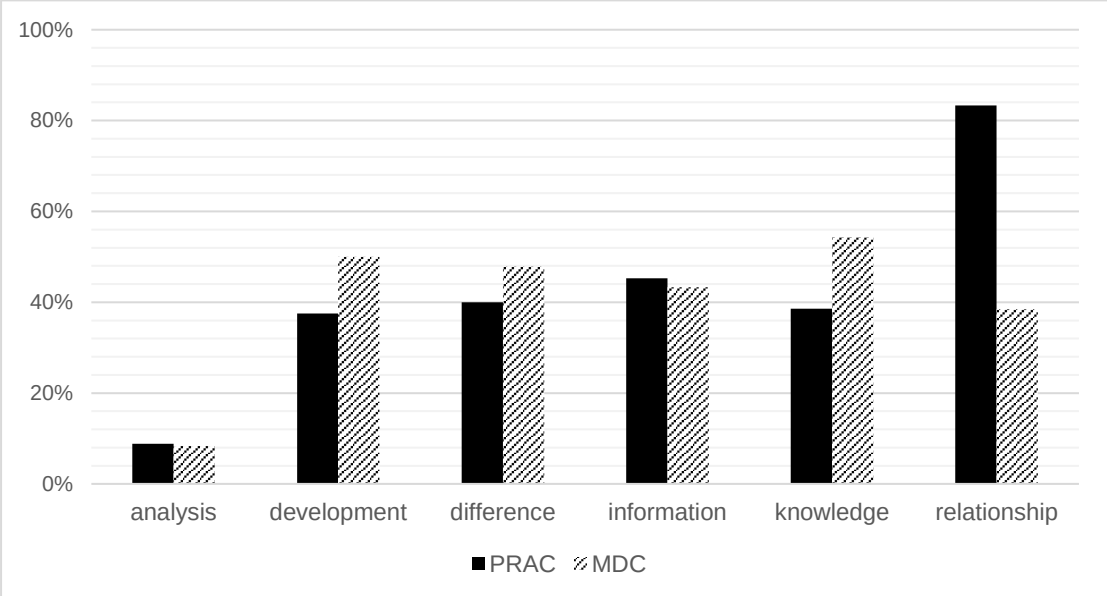


Figure 5.10. Proportion of clausal postmodification by finite relative clauses for each noun.

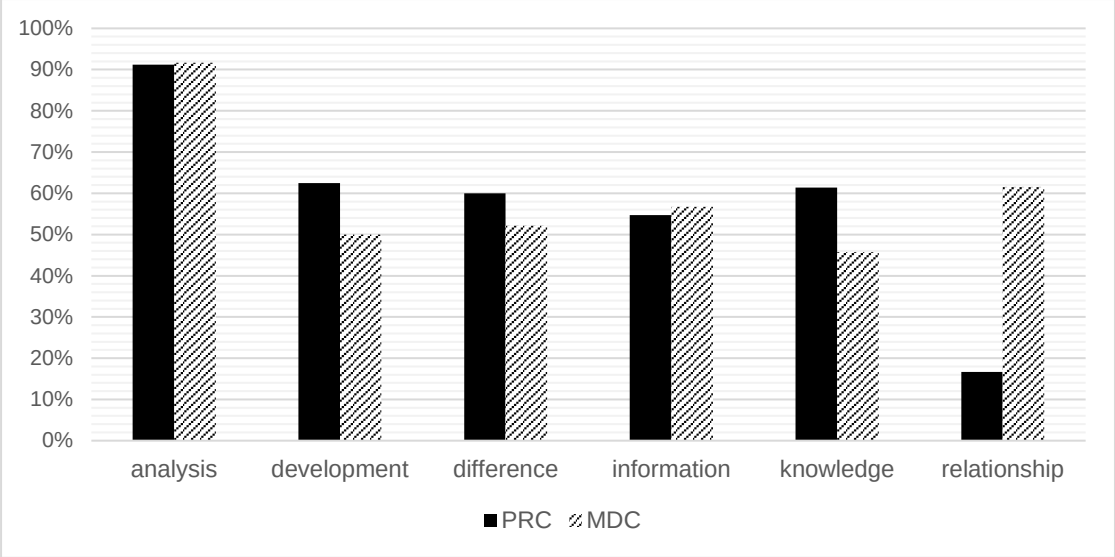


Figure 5.11. Proportion of clausal postmodification by non-finite clauses for each noun.

In summary, this section has examined the extent to which the selected NP samples from the two corpora are postmodified and how different types of postmodifiers are used in the NP samples. Analyses have been conducted for both the NP samples as a whole and for NPs headed by each individual head noun. A number of findings have emerged: 1) the two NP samples equally postmodified (42%); 2) both student writers and expert writers predominantly use prepositional phrases as opposed to clausal structures to postmodify nouns, although they are found to use slightly higher clausal postmodification; 3) in using PP postmodification, student writers use more *of*-PPs and fewer PPs headed by prepositions other than of; and 4) student writers use more finite and less non-finite clausal postmodification. However, it has also been found that the variations between the two collective NP samples do not always represent variations manifest in NPs with individual head nouns, indicating student writers' differential colligational development with respect to opting for postmodifier types.

Now that student writers do not differ much from expert writers in using preposition phrases as the predominant postmodifier type, it would be necessary to further examine the structural complexity emanating from PP postmodification—multiple postmodification, the most complex grammatical feature in Biber et al.'s hypothetical framework.

5.2.3. Multiple Postmodification in the Noun Phrase

The examination of multiple postmodification is based on a sample of complex noun phrases (CNPs) expanded from three head nouns (*understanding*, *analysis*, and *lack*), whose first level of postmodifier is an *of*-preposition phrase. Table 5.1 lists the number of occurrences of such CNPs. Two specific dimensions of multiple postmodification were examined: 1) length (i.e. number of words) of the complete postmodification and 2) depth of multiple postmodification, referring to the number of levels of multiple postmodification of the selected head nouns and the proportion of multiple-postmodified CNPs at each level (see Section 3.5.2.5 for details of analysis).

TABLE 5.1

Occurrences of Multi-Postmodified CNPs in the Two Corpora

	MDC	PRAC
understanding	232	251
analysis	272	275
lack	152	224

Before presenting the inferential statistical results, it would be interesting to revisit the two extended extracts used in 5.2.1 for the present purpose, re-presented as

Ex. 1’ and 2’. While Ex. 1 and 2 were highlighted for nouns with at least one postmodifier, highlighted in 1’ and 2’ below are complex noun phrases with more than one postmodifier:

(1’) ***The distinctiveness of writing item of CET 4****, the national foreign language examinations for professional advancement, is also worth discussing. Firstly, it is the writing for examination, which means the readers of the students’ writings are raters who are, in most cases, advanced writers. Secondly, it is designed with controlled length on essay (about 120 words) and on time allocation (Students are supposed to finish their composition within 30 minutes). Thirdly, different types of writing-task do pose different levels of difficulties for learners. The task types of CET 4 writing are mainly discursive (explanatory/argumentative) and these tasks involve* ***high levels of idiomatic, phrasal verb knowledge plus the ability to generate complex sentences****. The above mentioned factors, with the complexity of writing itself, make it more difficult for* ***students who could be regarded as lower-intermediate learners from vocational institutes****. Hereby, it makes sense that students take the shortcut by copying the model sentences.* [152 words] (Text 61, MDC)

(2’) *In addition, owing to* ***the exodus of emigrants into English-speaking countries (the United States in particular)****, since the early 1980s, there has been* ***a subsequent return of overseas Chinese whose Chinese language has been noticeably influenced by the English language not only in vocabulary and grammar but in pragmatics****. They have retained remnants of Anglo-American influence not only when conversing with each other, but also in their communication with local residents, since this supports* ***their reputation as***

> ***foreign-language-speaking cosmopolitans (for the most part accompanied by certain wealth)***. *Therefore,* ***the influence of English on Chinese*** *has progressed further in extent and increased in intensity at the current stage of language development. Not only has it accumulated a large body of information on lexical borrowing, morphological transference and syntactic variation, but as we have seen in this study, also influenced the pragmatics of the Chinese language. A good case in point is* ***the increasing use of appreciation CR strategy in Chinese speakers***. *Yuan (2002) also offers* ***discussions about the change of culture in China***, *arguing that the social changes are the reasons for her subjects to adopt more acceptance, but less non-acceptance strategies than previous studies on Chinese CRs.* [197 words] (Text 16, PRAC).

The highlight visualisations easily demonstrate a greater representation of CNPs with multiple postmodification in the PRAC than in the MDC. This forms a contrast with the overall picture of the extent to which nouns tend to be postmodified, for which the two corpora show a striking similarity. While examining the tendency for nouns to be postmodified does not reveal any difference in using postmodification between the two corpora, the use of multiple postmodification in constructing CNPs may do. Therefore, it would be interesting to see if the quantitative results to be presented supports this preliminary observation.

5.2.3.1. Length of Postmodification

First, the mean lengths of postmodification for the three sets of CNPs respectively expanded from each head noun in the two corpora were subjected to independent samples *t*-tests in order to test if mean length differences between the two

corpora are statistically significant.[33] Descriptive and statistical results are listed in Table 5.2.

TABLE 5.2

Comparison of Length of Postmodification in the Two Corpora

	MDC	PRAC	Cohen's *d*
understanding	5.59**	7.10	-0.355
analysis	5.93*	7.04	-0.269
lack	5.59*	6.71	-0.269

*. $p < .01$

**. $p < .001$

As shown in Table 5.2, a significant difference in mean postmodification length for CNPs with each head noun exists between the two corpora. This result is compatible with a previous finding: there is a significant difference in the mean length of clause between the two corpora (see 4.2.2). But as mean length of postmodification only shows a centralised tendency for the selected CNP samples, it does not address cases where less complex postmodification is used. The following examples illustrate the use of extremely short (2-3 words) and extremely lengthy (as long as 20 words) postmodification in both corpora, when CNPs have different information statuses across discourse: [34]

(5) *At appropriate points, they asked questions that explored the students' understanding **of plagiarism** (including inviting a definition of plagiarism) and*

[33] The results shown in this table and their discussions have been published in Liu and Li (2016).
[34] These examples and their discussion have been published in Liu and Li (2016).

questions that were derived from analyzing textual data. [2 words) (Text 118, PRAC)

(6) *A pun involves two meanings of a word, and as we illustrated in the previous part, the understanding* ***of the puns*** *requires the connection of two mental spaces.* [3 words) (Text 57, MDC).

(7) *The analysis of the student GBM provides an initial understanding* ***of the extent to which the students believed in the existence of FLA in foreign language learning***. [17 words) (Text 46, PRAC)

(8) *The O'Malley and Chamot's (1990) theory provides the learners with a deeper understanding* ***of metacognitive strategies as a kind of higher executive strategies than the other two kinds of strategies*** *to plan, monitor and evaluate the whole vocabulary learning process and the six kinds of subcategories of metacognitive strategies, i.e., advance organizers, selective attention, self-management, advance preparation, self-monitoring, self-evaluation, checked and explored in the present study.* [17 words) (Text 67, MDC).

Thus, it is necessary to examine the extent to which more complex postmodification is used to complement the comparison of mean length differences.

For reasons of pure comparison, an arbitrarily determined cut-off of eight words was set to identify postmodification longer than eight words as "more complex", and their proportional representations in the two corpora were computed. Table 5.3 shows the proportion of more complex postmodification in the two corpora.[35] As can be seen, there is a smaller representation of more complex postmodification in the MDC than in the PRAC. While more complex postmodification in the MDC has a representation ranging from 20 per cent to 26 per cent, the range is 35 per cent to 37 per cent in the

[35] The results shown in this table and their discussions have been published in Liu and Li (2016).

PRAC. It can thus be said that not only is postmodification shorter in the student corpus, student writers also use a smaller amount of more complex postmodification.

TABLE 5.3

Representation (%) of More Complex Postmodification in the Two Corpora

	MDC	PRAC
understanding	21%	35%
analysis	26%	37%
lack	20%	35%

5.2.3.2. Depth of Postmodification

This study calculated the proportion of multi-postmodified CNPs of different complexity levels distinguished by the number of consecutive postmodifiers. This information is presented in Table 5.4, with the level numbers indicating the numbers of consecutive postmodifiers. It can be seen that both corpora include CNPs with four consecutive postmodifiers for all three head nouns; and both corpora have CNPs headed by *analysis* and *lack* with five consecutive postmodifiers. However, CNPs with five or more consecutive postmodifiers are rare in both corpora, making their comparisons not necessary.

TABLE 5.4

Proportion of Multiple Postmodifiers

		understanding		*analysis*		*lack*	
		PRAC	MDC	PRAC	MDC	PRAC	MDC
Max. no. of levels		5	4	6	5	8	5
Instances of CNPs		251	232	275	272	224	152
Level 2	Instances	116	81	129	113	119	67
	Proportion	46%	35%	47%	42%	53%	44%
Level 3	Instances	31	16	42	38	55	25
	Proportion	12%	7%	15%	14%	25%	16%
Level 4	Instances	8	4	11	11	14	6
	Proportion	3%	2%	4%	4%	6%	4%
Level 5	Instances	2	--	1	2	6	2
	Proportion	0.80%	--	0.40%	0.70%	3%	1%

As can be seen from Table 5.4, for CNPs expanded from all three head nouns, the PRAC sees a greater proportion of CNPs with two, three, and four consecutive postmodifiers in comparison with MDC.[36] An exception is CNPs expanded from *analysis* with three and four consecutive postmodifiers, which show an identical or almost identical percentage in the two corpora. For instance, 46% of CNPs expanded from *understanding* in the PRAC have at least two postmodifiers versus merely 35% in the MDC. Similarly, 12% of the same CNPs in the PRAC have three postmodifiers compared with 7% in the MDC. Another observation is that CNPs with two or three postmodifiers exhibit a larger gap than those with four postmodifiers. In general,

[36] The results shown in this table and their discussions have been published in Liu and Li (2016).

multiple postmodification in the student corpus has shown weaker depth compared with the expert corpus, with the difference mainly attributable to CNPs with two and three consecutive postmodifiers, and those with four postmodifiers only showing a negligible difference.

As previously presented, CNPs expanded from *analysis* have shown a smaller difference in postmodification depth between the two corpora. This could be explained by the observation that students, as do expert writers, employ the bigram *analysis of* to describe their own studies in great length. Thus, students' intimacy with their own research (e.g. the object, method, and aim of analysis) has a contribution to make for the production of detailed CNPs expanded from *analysis*, such as the following example:

> (9) *Item 16 implies that the analysis of the types of one's learning tasks and decision on the corresponding learning strategies are also significant in language learning, and this can accelerate their English achievements.* (Text 21, MDC)

The examination of the complexity of NP postmodification thus far has shown weaker length and depth in the student corpus. The finding is in accordance with the results of syntactic complexity measures related to noun phrase complexity: mean length of clause, complex nominals per clause, and complex nominals per T-unit (see 4.2.2). Regarding the depth of multiple postmodification, two observations can be made: 1) expert writers do not usually use CNPs with three or more postmodifiers, with half of the selected CNPs in the expert corpus having two postmodifiers; 2) students trail experts mainly in CNPs with two and three consecutive postmodifiers, with CNPs with four or more postmodifiers being a rare phenomenon. This feature of multiple postmodification in the CNP may indicate a natural tendency of professional academic writers to use

consecutive postmodifiers that is conditioned by such cognitive factors as short-term memory, ideal boundary of intonation unit, and vicarious perception of reading difficulty.

Useful implications for EAP instruction can be drawn from these findings. Although a prominent feature of academic discourse, expert writers' use multiple postmodification should not be billed as extremely complex. Instructional effort, therefore, needs to be made to help students elaborate information by effectively using two or three consecutive postmodifiers.

5.2.3.3. Textual Evidence of Quantitative Findings

This section presents textual analysis of student writing to support quantitative results of syntactic complexity measures and postmodification complexity. Differences in the clause length and NP complexity in elaborating information in student writing can be demonstrated by the following examples found in the introduction of an MA dissertation and a research article: [37]

(10) ***The information expressed in either speaking or writing*** *should be clear enough to be understood by certain structure. It is not only* ***the requirement of everyday communication****, but also* ***the requirement of any other written text in books, newspapers, magazines and so on****. This is because a special property, which is called 'coherence', is lying in language. Coherence is an important notion in discourse analysis, and it exists everywhere in our daily life. It is* ***the semantic or functional relation that implies in the deep structure of discourse*** *that produces coherence.* [91 words] (Text 42, MDC)

(11) *Research has shown just how significant* ***proficiency in the main language***

[37] These examples and their discussion have been published in Liu and Li (2016).

of the destination country *is to t***he economic integration *of immigrants in Canada and the United States***. *Specifically,* ***an increase *in proficiency in this main language*** *enhances* ***the effects *of earnings from education and the labour experience of the country of origin***. *In the United States,* ***immigrants who are proficient in English*** *have earnings approximately 17% higher than* ***immigrants *with limited English proficiency***. ***A report *by Statistics Canada using data from the 2006 census*** *highlights* ***the economic benefits *of speaking only English or French in the workplace***. [98 words] (Text 6, PRAC)

Ex. 10 includes nine clauses consisting of 91 words and Ex. 11 contains six clauses with 98 words, making Ex. 10 less complex than Ex. 11 in mean clause length: 10 versus 16 words per clause. What makes the clauses in Ex. 11 contain more information is its more intense use of noun phrases with multiple postmodification than Ex. 10: eight CNPs in six clauses as opposed to just four in nine. In other words, there are an average of 1.3 CNPs per clause in Ex. 11, but only about 0.5 in Ex. 10. Such a large numerical difference in the density of complex noun phrases and their length in the two examples is easily discernible from text enhancement: CNPs are marked in bold and postmodification underlined. Furthermore, Ex.11 has more CNPs with multiple postmodifiers than Ex. 10 (5 versus 2). Also striking is that four of the eight CNPs in Ex. 11 fill the grammatical position of subject whereas there is only one CNP subject in Ex. 10. This finding resonates with Vande Kopple's (1994, p. 534) observation that "grammatical subjects in scientific discourse are markedly long" and Master's (1991) finding that inanimate subjects with active verbs are frequent in academic writing and this structure presents is difficult for EFL writers. In contrast, most of the subjects in Ex. 10 (60%) are deictic words (i.e. *which*, *it*, and *this*), a major characteristic of

context-dependent genres requiring much reader/listener involvement, such as informal conversations (Biber, 1988).

A further note is that the lack of apparent syntactic complexity in Ex. 10 is also derived from using coordinate clause frames (i.e. *it is not only… but also…* and *Coherence is… and it…*), creating a strong pattern of parallelism as a text-unfolding rhetorical device. However, if parallel constructions are frequently used, they "would stand in the way of clear communication" (Kaplan, 1966, p. 8). Parallelism in fact underlies Ex. 10 as the subjects of Sentences 2 to 4 are merely deictic repetitions of statements expressed in the previous sentence. While it apparently observes the principle of information structure from given to new (Quirk et al., 1985), simple repetitions of given information without specifying the context in question may lead to frequent logic breakups and reading monotony. By contrast, the CNP subjects in Ex. 11 effectively builds upon and expands upon given information.

When the two examples are more carefully observed, it is possible that the student writers use loose grammatical structures such as those in Ex. 10 because they fail to construct strong logical relations between statements. The use of parallel constructions analysed above blurs logical connections among statements and their identification is only left to the reader's own decision. Building such connections requires finite clauses to be compressed to nominal structures which can then be combined (see Tannen, 1982, p. 39 ff.). The important mechanism for meaningfully connecting nominalised constituents is hinged on effectively selecting predicative vocabulary, typically through lexical verbs or linking verbs plus complements. This mechanism can be clearly discerned in Ex. 11, especially in frequently using lexical verbs to link nominal structures, including *show*, *enhance*, and *highlight*. Although grammatically simple, this simple subject-verb-object sequence with complex nominal

participants nonetheless presses the writer to identify the logical links among nominal participants and to retrieve their lexical knowledge to explicate the relations. In this sense, the fact that Ex. 10 almost has none of this discourse feature may indicate that the student is purposely dodging the cognitive challenge of synthesising and integrating information or simply fails to detect the implicit relations among statements. Both ways, however, there is the obvious need for overt teaching on this syntactic strategy.

However, one needs to recall that student writers also use highly complex noun phrases (e.g. CNPs expanded from *analysis*), notably when students try to describe in detail the research methodology of their own research. It is also attested that some EFL writers have been already skilful in using sophisticated noun phrases with rich predicative vocabulary as the major discourse unfolding strategy, exhibiting little difference from expert writers, such as:

> (12) *Vocabulary and grammar teaching, along with phonetics teaching, constitute the core of English language teaching in junior middle schools in China. Despite the fact that English teachers have long realized the importance of teaching vocabulary and grammar in an effective way, a large number of them continue to teach them in a rather traditional way, that is, filling their classrooms with boring illustrations and drills. Traditional teaching methodologies still exert influence on English teachers, resulting in their mindset. For example, many teachers still attach far greater importance to the role of grammar and vocabulary in translating texts than to the context in which language is used. It is characteristic of the grammar-translation method which used to be predominant in China. Some others put too much emphasis on endless but frustrating error correction in listening and speaking, which is typical of the audiolingual method.* (Text 23, MDC)

However, the problem is that expert-like noun phrase complexity does not occur stably in EFL writing, some showing a much more sophisticated manipulation than others. At the same time, it needs to be noted that an important sign of grammatical maturity is reflected in the writer's ability to vary sentence length. Sometimes short sentences can function to draw readers' attention, such as the following topic sentence taken from the expert corpus:

> (13) ***We will start the discussion with perhaps the most intriguing item: the***. *The difficulty Chinese learners have with the English article system has long been noted (Cai & Wu, 2006), with both overspecification (e.g., the for a/an or Ø) and underspecification (e.g., a/an or Ø for the) errors reported (Chuang and Nesi, 2006; Díez-Bedmar and Papp, 2008; Master, 1995). Various frameworks have also been proposed for teaching articles (e.g., Master, 1990). The initial findings from our keywords analysis show that the definite article constitutes 6.3% of the tokens in EXJA but 7.4% CAWE…* (Text 64, PRAC).

The implication for EAP instruction is that although complex noun phrases dominate expert writing, this feature cannot be rigidly followed as invariable for text development; factors such as stylistic diversity and rhetorical functional also need to be considered.

This section has examined the complexity of multiple postmodification from the perspectives of its length and depth. It has been found that student writers produce shorter and fewer levels of consecutive postmodification than expert writers. However, multiple postmodification in expert writing should not necessarily entail extreme complexity: most cases of CNPs in the expert corpus have two or three consecutive postmodifiers.

Furthermore, the text-based analysis of student writing has provided concrete evidence for the quantitative findings of shorter clause length, smaller complex nominal ratio, and underdeveloped postmodification in student writing. The textual analysis also shows how premature use of complex noun phrases results in a discourse flow that lacks cohesion, coherence, and authorial evaluation. Although it cannot be said that noun phrase complexity guarantees good coherence and writing quality, it would be safer to say that complex noun phrases, if effectively arranged, can greatly enhance writing quality. This is consistent with a finding from McNamara et al (2009), that is syntactic complexity measured by the length of unit before the predicate verb is one of three best predictors of first language writing quality (together with two measures of lexical richness). In this regard, the following section examines in detail how complex noun phrases are used as clause subjects in student writing.

5.2.4. Complex Noun Phrases as Clause Subjects

This section examines the extent to which student writers use complex noun phrases as clause subjects, a distinctive feature of academic discourse (Vande Kopple, 1994). The same sets of CNP samples used for 5.2.3 were employed as the data for this analysis. CNPs expanded from each head noun were coded for their occurrences as subjects of finite clauses. The extent to which CNPs serve as clause subjects was calculated as the proportion of subject-filling CNPs to all occurrences of sampled CNPs in each corpus; meanwhile, the mean lengths of subject CNPs in the two corpora were compared for statistical significance. Additionally, the proportion of longer CNP subjects (those over 10 words long) was also examined to further understand differences (if any) in the complexity of CNP subjects (see Section 3.5.2.6 for detail of analysis).

5.2.4.1. Proportion and Length

Table 5.5 presents the proportions of CNPs expanded from the three head nouns serving as clause subjects. It can be seen that the CNP subjects of all three head nouns appear more frequently in the PRAC than in the MDC. However, of note is the fact that while the gaps between the two corpora for CNPs expanded from *understanding* and *lack* are rather moderate (respectively 1% and 3%), that for CNPs from *analysis* is a most striking one—21.3%. That is to say, expert writers tend to use *analysis*-CNPs at the start of clause much more frequently than student writers. An analysis of the concordance lines containing these CNPs reveals that many of them play the role of an inanimate agent in the clause and would typically take reporting verbs, e.g. *show*, *reveal,* and *indicate*:

> (14) *An analysis of the DCT-1 utterances* ***showed*** *that the learners as a group used the target request-making forms in 35 of the 120 utterances, leading to an appropriateness rate of 29.16%.* (Text, 104, PRAC)
>
> (15) *An analysis of the first item cluster* ***indicates*** *that the words in and find, which had been found to misfit the Rasch model under item analysis in Winsteps, clustered under a single dimension regardless of the form of knowledge tested*. (Text 25, PRAC)

This finding also seems to suggest that student writers in general may have acquired a somewhat different schema of the functional and semantic properties of CNPs headed by certain abstract nouns, for example *analysis* in the present case. As introduced in 2.4.2.3, the use of inanimate/abstract subjects is a distinct syntactic feature of academic discourse (Biber et al., 1999; Master, 1991; Vande Kopple, 1994), whether taking verbs typical of human action or those expressing abstract relationship/existence. In this sense, it may also be argued that student writers have not developed a full-fledged awareness

of abstract nominals serving as clause subjects to the extent that has been reached by expert writers, although their texts have already exhibited a relatively high representation of this syntactic feature and in some cases are nearly on a par with published texts.

TABLE 5.5

Proportions of CNPs as Clause Subjects

	MDC	PRAC
understanding	13% (30)	14% (35)
analysis	22% (60)	43% (118)
lack	18% (27)	21% (47)

Note. The figures in brackets indicate the number of CNP subjects identified.

Although student writers use CNP subjects as frequently as expert writers for CNPs headed by *understanding* and *lack*, when the comparison focuses on the length of CNP subjects a much sharper contrast appears. As described in 3.5.2.6, for pure comparison's purpose, an arbitrary cut-off point of 10 words was set to identify CNPs ≥ 10 words as greater-length productions and to discover their presence in the selected CNP sample. As Table 5.6 shows, there are far more occurrences of greater-length CNPs with all three head nouns in the PRAC than in the MDC: those in the PRAC account for 20%–33% of all instances of CNPs; in contrast, those in the MDC expanded from *understanding* and *lack* have little or none representation at all (3.7% and zero respectively).

TABLE 5.6

Proportions of CNP Subjects of ≥ 10 Words

	MDC	PRAC
understanding	4% (1)	20% (6)
analysis	17% (10)	33% (40)
lack	0 (0)	30% (14)

Note. The figures in brackets indicate the number of CNP subjects identified.

A more in-depth analysis of students' use of CNP subjects can be obtained from comparing the mean lengths of subject CNPs in the two corpora. Table 5.7 presents a significant difference in the mean length of CNP subjects expanded from all three head nouns between the two corpora, with the gap ranging from approx. 2 words (*understanding* and *lack*) to around 1 word (*analysis*).

TABLE 5.7

Length of Subject-CNPs in the Two Corpora

	MDC (*SD*)	PRAC (*SD*)	Sig.	Cohen's *d*
understanding	6.444 (1.888)	8.300 (4.419)	.042*	-.54
analysis	7.328 (2.612)	8.336 (3.607)	.016*	-.36
lack	5.929 (1.942)	8.000 (3.887)	.002*	-.69

*. Difference is significant at the $p < .05$ level.

The effect size Cohen's *d* values representing the strengths of differences (-.54, -.36, and -.69 respectively) shown in Table 5.6 range from lower-medium (-.36) to higher-medium (-.69) effect sizes. These fairly large strengths of significant differences

in the length of CNP subjects could be explained by the differing standard deviations (*SD*) of the two datasets: the smaller *SD* for the MDC dataset indicates that CNP subject lengths in the student corpus tend to centre around the mean score while the larger *SD* for the PRAC dataset signifies that the CNP subject lengths in the published corpus are distributed in a manner more dispersed from the mean. That is to say, both shorter and longer CNP subjects exist in the expert corpus to a greater extent than they do in the student corpus. While producing of shorter-length CNP subjects is natural and legitimate as a result of discursive constraints (e.g. *a better understanding of this problem* where *this problem* has been adequately described in previous discourse), using greater-length CNP subjects represents the writer's greater textual awareness to integrate and synthesise information mentioned in previous discourse.

The above findings suggest that student writers are even less prepared for producing complex nominal structures to initiate a statement compared with their use for other syntactic functions, e.g. objects and complements. Psycholinguistic research into the relationship between syntactic complexity and speech production has found that syntactically complex subjects tend to cause speakers to pause at the subject-verb phrase boundary, due to prolonged information retrieval from short-term memory (Ferreira, 1991). Such a psycholinguistic challenge may partly explain what holds the students back in producing complex subject-CNPs.

5.2.4.2. Discourse Analysis of CNP Subjects

The discourse analysis of students' use of CNP subjects is based on the three sources of pressure prompting research scientists to use extended nominal subjects proposed by Vande Kopple (1994), as discussed in 2.4.2.3. These are the pressure to be precise, the pressure to be concise and economical with words, and the pressure to be

efficient and progressive in constructing claims; the last pressure is actually derived from the first two and from the broader socio-semiotic situation in which writing is situated.

It could be contended that student writers, whose texts have seen markedly low representation of CNP subjects, may not have been faced with these pressures in writing up their dissertations and more probably that they may not have been made aware of the discursive functions of CNP subjects. For example, the fact that CNP subjects used in the MDC generally include fewer levels of postmodification and that the length of each postmodifying constituent is also much shorter in the student corpus make a convincing case that student writers are not informative enough in producing a subject that can adequately relate previous segments of discourse. Figure 5.12 presents some typical examples of CNP subjects expanded from *understanding* in the student corpus. As the use (and non-use) of the indefinite article *a/an* and the premodifying adjective *better* implies, all instances of the examples attempt to provide specific and new characterisations for the head noun *understanding*, but obviously the information contained in each instance is of a general category and lacks in the necessary substance of specificity.

Theoretically, through this research, the essence of the uses of the simple present tense can be uncovered, so **a better understanding of the simple present tense** can be achieved.
According to Hu & Gao (1997), **a better understanding of competence** enables an individual to have a better acceptance of self, others and the nature, thereby having a better insight into reality.
And the recognition and **understanding of new things** are based on those primary and basic experiences.
An understanding of other cultures and the ability to communicate effectively and work with people from these cultures is absolutely necessary for those who work internationally.
So **an understanding of attribution theory** will help to identify what factors are harmful or beneficial, to find out reasons for lack of persistence and to understand in what way to attribute can we improve our learning achievements
In fact, the relationship between culture and language can be summarized to be that **an understanding of a language** is impossible without constantly relating it to the culture in which it is operative.
An accurate understanding of deixis not only can help the speaker use it more properly but also can help the receiver interpret the speaker's intention more accurately.

Figure 5.12. Examples of understanding-CNPs as clause subjects (highlighted in bold) in the MDC.

Being general and non-specific, student writers do not encapsulate in the CNP subjects the entities and processes relevant to the "understanding" under discussion and fail to distinguish their own studies from others by giving careful and detailed specifications to the themes in question. Another functional dimension to the use of long subjects for the sake of precision, according to Vande Kopple, arises from researchers' concern about making their claims less vulnerable to skepticism and challenge. As Vande Kopple (1994) writes,

> It is as if they [researchers] were stating something like this: "What we are saying is to be taken as true in these particular respects of these particular individuals on this particular day after being treated in this particular way with this particular drug". (p. 548)

Therefore, this near-obsessive pursuit of precision evidently requires a noun phrase structure with ample premodification and postmodification. Apparently, relatively pressure-free student writers do not seem to have to be concerned about this exigency

of being exact. In contrast, as can be seen from Figure 5.13, expert writers attempting to introduce particular types of "understanding" into the discourse (marked by the indefinite article *a*/*an*) more often than not provide rather elaborated characterisation to make the distinguishing aspects of their own studies stand out and to carefully contextualise the entities and processes necessarily concerned.

Samraj's work shows that consideration of the disciplinary context alone is not sufficient and suggests that **an understanding of the 'layers of context' and their influence on student roles and texts produced** should constitute an important part of the specificity we want to teach
An understanding of the classification and distribution of genre families is furthermore essential for effective investigation of the BAWE corpus by teachers, students, textbook and other materials writers, and researchers.
In this regard, **a better understanding of learner-generated noticing processes** seems timely and worthwhile.
A clear understanding of what constitutes acceptable imitation and borrowing for Chinese students can shed further light on their knowledge of and attitudes toward plagiarism.
Varghese, Morgan, Johnston, and Johnson (2005) maintain that **a comprehensive understanding of teachers, teaching, and teacher education** requires attention to both identity-in-discourse and identity-in-practice.
A deeper understanding of the motivations of its minorities is essential to meeting their needs
A deeper understanding of learners' internally generated noticing process should have important implications for both researchers and practitioners in their efforts to increase the likelihood of 'matching' the instructional intervention (i.e. externally created salience) with the
The understanding of language learning processes has been a focus in longer courses (Ellis, 2009, 2010; Peacock, 2009).
The implication is that students are likely to be confused if **their understanding of good paraphrasing** is different from that of their professors.
This understanding of writing strategies is, however, at odds with the recent development of cognition research in at least three ways.
This understanding of language education is based on the premises of postmodernity, a period that, according to Hall (2006), emerges from the decentralization of the modern way of life.

Figure 5.13. Examples of *understanding*-CNPs as clause subjects in the PRAC.

For example, instead of a generalised *a deeper understanding of learners' learning process*, quite typical of CNP subjects in students' texts (see Figure 5.12), what is phrased in the expert corpus turns out to be *a deeper understanding of learners' internally generated noticing process,* which is precise enough to distinguish from

other kinds of "learning process". Thus, results arising out of "a deeper understanding of learners' internally generated noticing process" can be solidly contextualised and secured from criticism due to otherwise vague characterisation.

It needs to be noted that when expert writers do use CNP subjects with simple and generalised information, it is usually known information that they are bringing up, as signalled by the use of the definite article *the*, demonstratives (e.g. *this*), and pronouns (e.g. *their*), as shown by the last few examples in Figure 5.13; old information, as discussed above, generally does not require elaborated characterisation. But when the flow of discourse requires the subject to bring together a number of previously mentioned elements or those from the broader exophoric research context, CNP subjects with definite articles or demonstratives can also be lengthy, as shown by the following example from the expert corpus:

> (16) ***The lack of transparency (i.e., explicit discussion within the institution) that surrounded many of these issues*** *was a central factor in the ambiguities and dilemmas that CETs experienced in relation to doing research*. (Text 21, PRAC)

It can be imagined that the entities and processes included in this noun phrase (i.e. *lack*, *transparency*, *discussion*, *institution*, and *these issues*) have already been taken up in previous segments of discourse and the CNP functions to pack them up in an efficient and precise manner to pave way for introducing the newer message in the predicate.

The above discussion has attempted to reflect on students' markedly short and unelaborated CNP subjects mainly from a discourse-oriented information structuring point of view. Students' seeming lack of syntactic sophistication to use CNP subjects to pack information either for introducing new elements into the discourse or for summarising those previously mentioned could be considered as implicitly grounded

in their underdeveloped awareness of discourse-oriented principles of information structuring in academic discourse. It might also be possible that students have generally resorted to the end-weight/focus principles of information processing and have therefore intentionally rendered subject-position elements less elaborated. However, extrapolations as such are only based on examination of texts and should be further corroborated by insightful ethnography of students' writing processes and their introspection into the factors leading up to the lexicogrammtical decisions they make.

5.2.5. Predicate Vocabulary of CNP Subjects

In order to better understand how CNP subjects are semantically represented in the two corpora, the study examined the transitivity patterns and lexical richness of the predicate vocabulary taken by these CNPs. Tables 5.9-5.11 list all the lexical items used as predicate vocabulary for the three sets of CNP subjects in the two corpora, together with their number of occurrences (see Section 3.5.2.7 for analytical procedure). Table 5.8 presents information on the transitivity patterns of these lexical items in the two corpora. This section starts with a brief look at the transitivity patterns of the lexical items and then goes on to examine their lexical richness in association with results on transitivity.

TABLE 5.8

Transitivity Patterns of Lexical Items for CNP Subjects' Predicate vocabulary and Their No. of Types

		understanding		*analysis*		*lack*	
		PRAC	MDC	PRAC	MDC	PRAC	MDC
Transitive verbs	Monotransitive	10	3	23	16	16	6
	Ditransitive	0	0	0	0	2	2
	Complex transitive	0	3	1	1	6	2
	Total transitive	10	6	24	17	24	10
Intransitive verbs	Used alone	3	1	3	4	4	1
	Used in phrases	1	0	5	5	3	4
	Total intransitive	4	1	8	9	7	5
Copular verbs	Copula + nominal	1	1	2	2	3	3
	Copula + adjective phrase + complement (optional)	10	6	7	3	2	3
	Copula + preposition phrase	1	1	0	0	1	1
	Total copular	12	8	9	5	7	7
Total types		26	15	41	31	38	22

TABLE 5.9

Types of Predicative Items for the *understanding*-headed CNPs in the Two Corpora

	MDC (30)		PRAC (35)	
#	Verb phrase types	Occurrences	Verb phrase types	Occurrences
1	**help**	4	facilitate	2
2	**require**	4	constitute	1
3	enable	1	shed light on	1
4	prepare…for…	1	**require**	2
5	do	1	lead to	2
6	arouse	1	have implications for	1
7	mean	1	provide	1
8	determine	1	**help**	1
9	**be based on**	2	include	1
10	be in accordance with	1	come from	1
11	be rooted in	1	complement	1
12	be associated with	1	consolidate	1
13	be impossible	1	be essential for/to	2
14	be necessary	2	be crucial to	1
15	be possible	1	be integral to	1
16	be a process	1	be ahead of	1
17			be tranferable to	1
18			be different from	1
19			be at odds with	1
20			**be based on**	1
21			be limited to	1
22			seem timely and worthwhile	1
23			improve	1
24			develop	2
25			remain limited to	1
26			be a focus	1

Note. Items in bold occur in both corpora. The figure in brackets indicate the instances of CNP subjects.

TABLE 5.10

Types of Predicative Items for the *analysis*-headed CNPs in the Two Corpora

	MDC (60)		PRAC (118)	
#	Verb phrase types	Occurrences	Verb phrase types	Occurrences
1	**show**	3 + 2 (that clause)	**reveal**	15 + 13 (that clause)
2	**provide**	2	find	3 +2 (that clause)
3	constitute	2	**provide**	3
4	help	1	identify	2
5	depict	1	indicate	2 + 5 (that clause)
6	imply	1	suggest	2 +5 (that clause)
7	**reveal**	3	aim at	1
8	enable	1	assist	1
9	**focus on**	1	refute	1
10	represent	1	explore	1
11	lead to	1	draw on	1
12	involve	1	corroborate	1
13	deal with	1	elaborate on	1
14	offer	1	reflect	1
15	explain	1	allow	1
16	consist of	2	detect	1
17	consider	1	account for	1
18	indicate	1 (that clause)	include	1
19	have	1	**focus on**	1
20	begin (intr.)	1	warrant	1
21	develop (intr.)	3	uncover	1
22	advance (intr.)	1	demonstrate	4 (that clause)
23	prevail (intr.)	1	**show**	14 (that clause)
24	come into being (intr.)	1	start (intr.)	1
25	**be based on**	3	operate (intr.)	1
26	be significant	1	originate (intr.)	1
27			be helpful for	1
28			**be based on**	2
29			be basic in	1
30			be biased towards	1
31			be necessary	1
32			be fruitful	1
33			be limited	1
34			be one of the…	1
35			be one facet	1

Note. Items in bold occur in both corpora. The figure in brackets indicate the instances of CNP subjects.

TABLE 5.11

Types of Predicative Items for the *lack*-headed CNPs in the Two Corpora

	MDC (27)		PRAC (47)	
#	Verb phrase types	Occurrences	Verb phrase types	Occurrences
1	**lead to**	4	constrain	1
2	**make**	2	**make**	2
3	influence	1	provide…with…	1
4	bring about	1	correspond to	1
5	cause	1	serve as…	1
6	yield	1	accentuate	1
7	**contribute to**	1	hinder	1
8	do harm to	1	subject…to…	1
9	stand as…	1	place…at…	1
10	account for	1	constitute	1
11	**result in**	1	hamper	1
12	reduce	1	restrict	1
13	prevent…from	1	impact (on)	2
14	**mean**	1 (that clause)	explain	1
15	imply	1 (that clause)	undermine	1
16	be due to	1	imply	1
17	be reasonable	1	have	1
18	be blocks	1	**contribute to**	2
19	be reason	1	raise	1
20	be fatal	1	limit	2
21	be problem	1	hold…back	1
22	be attributed to	1	call into question	1
23			**lead to**	3
24			**result in**	1
25			motivate	1
26			stop…from…	1
27			help (intr.)	1
28			combine (intr.)	1
29			apply (intr.)	1
30			**mean**	2 (that cluase)
31			indicate	2 (that cluase)
32			suggest	1 (that clause)
33			be linked with	1
34			be at the root of	1
35			be culprits	1
36			be factor	1
37			be difficulty	1
38			be noteworthy	1

Note. Items in bold occur in both corpora. The figure in brackets indicate the instances of CNP subjects.

5.2.5.1. Transitivity Patterns

As shown in Table 5.8, the lexical items predicating the CNP subjects used by both groups of writers cover almost all the transitivity types coded for analysis. For both corpora, the predominant transitivity type employed is the use of transitive verbs, followed by copular verbs and intransitive verbs. Although rare, student writers also use ditranstive and complex transitive verbs, to an extent that is similar to expert use. For CNPs with each head noun expert writers use a greater variety of lexical items. However, a closer examination of the three transitivity types indicates that it is mainly in the use of transitive verbs where student writers lag behind expert writers. This may be due to the potential semantic difficulty inherent in using transitive verbs, caused by the fact that at least one other argument is needed as the object (complement) of the transitive verb. In other words, the (largely abstract) CNP subjects have to come into semantic relationships with one or more other participants, which are most probably also abstract, such as:

> (17) …, *the lack of L2 linguistic proficiency might have* ***limited*** *their cognitive resources for idea production and* ***constrained*** *the effectiveness of their L-S in planning their writing*… (Text 92, PRAC)
>
> (The lexical items predicating the CNP subject *the lack of L2 linguistic proficiency* are highlighted in **bold** and the abstract entities complementing the predicative verbs are underlined.)

Making meaning in such a manner would no doubt present a huge challenge for the writer, explaining student writers' apparent avoidance in using transitive verbs with CNP subjects.

Another point to note is that the use of copular verbs constitutes the second major transitivity patter in both corpora. Here, the construction copula + adjective phrase + optional complement is shown to be the predominant usage, while copulas used with nominals (e.g. *…the analysis of RA macro structures was only once facet*) and preposition phrases (e.g. … *this understanding of writing strategies is at odds with the recent development of cognition research*…) are not as frequent. The former also sees a stronger representation in the expert corpus, particularly associated with CNP subjects headed by *understanding* and *analysis*. Both student and expert writers have used adjectives expressing description (e.g. *different*) and evaluation (e.g. *timely*) in commenting on the abstract subjects (cf. Farsi, 1968), functions similar to those performed by using verbs.

The fact that expert writers use more predicate adjectives suggests that they are more flexible in choosing from other lexical resources when there are not appropriate lexical verbs for use. In the case of CNP subjects with *understanding*, the expert corpus includes eight unique adjectives with preposition phrase complements while the student corpus only three. The items used by expert writers include *be essential for/to, be crucial to, be integral to, be ahead of, be transferable to, be different from*, *be based on,* and *be limited to*; those used by student writers are *be rooted in, be associated with,* and *be based on*. As can be seen, the three items used by student writers are all adjectivised past participles derived from their passive voice semantics, while expert writers use both common adjectives and participle-turned adjectives. Using such adjectives as *integral*, *ahead,* and *transferable* to predicate *understanding* helps enrich

and expand its semantic scope such that it is metaphorised to be temporally and spatially associated with other participants in the clause, i.e. the nominals in the complements of the adjectives. Therefore, student writers need to be encouraged to broaden their perspectives on what abstract concepts like *understanding* can **do** and can **be**.

The frequent use of copular verbs with abstract subjects, in most cases the copula BE, should be seen as an important meaning making and discourse creating device, despite previous research labelling the use of copular verbs and verbs from certain semantic categories as simplistic and typical of immature EFL writing (Hinkel, 2003). While inexperienced L2 writers may tend to overuse copular structures due to a lack of lexical verb repertoire, what really makes meaning appear immature should be concerned with inadequate complexity encoded in nominals and their semantic relationships, not so much attributable to the use of copular verbs alone.

5.2.5.2. Lexical Variation of Predicate Vocabulary

As can be seen from Tables 5.10-5.12, CNP subjects with each head noun are predicated by a greater variety of lexical items in the PRAC than in the MDC, indicating expert writers' greater lexical variation in the selection of predicative resources for abstract and complex subjects. Among CNPs with the three head nouns, those with *understanding* and *analysis* see a gap of 10 and 9 unique lexical items while those headed by *lack* witness the largest gap, with expert writers using 16 more unique items than student writers. Of note is that although there are relatively fewer instances of subject CNPs headed by *understanding* and *lack* than by *analysis*, expert writers still use a great variety of predicate vocabulary for the two sets of CNP subjects. Thus, expert writers not only use a greater proportion of CNP subjects than student writers, they also have a greater inventory of predicate vocabulary to assign meaning to these

abstract concepts. This makes CNP subjects used by expert writers, which are already grammatically highly complex, also exhibit a greater semantic variety due to the varied selection of predicate vocabulary.

The overall greater lexical variation shown in the selection of predicate vocabulary in the published corpus can be regarded as reflecting a broader spectrum of meanings assigned to the subject-CNPs, suggesting a wider range of semantic possibilities associated with the CNP subjects as expert writers see fit. Student writers' relatively limited repertoire of predicative lexical items not just reflects their underdeveloped productive vocabulary knowledge (Read, 2000) but may also indicate that they have a limited understanding of what abstract processes and entities can "do" and can "be" and the possible relationships they enter into with other abstract existences. Alternatively, it is possible that students do have a perception of the roles abstract entities and processes can play but it is linguistically so fuzzy and elusive that they lack appropriate L2 lexical resources to articulate them. From another perspective, verbs are prototypically used with animate subjects in speech, most notably with nouns referring to people and objects, the tangible and concrete property of which makes the production of an utterance natural and congruent, hence cognitively easier.

By contrast, it would be necessarily more challenging to think of inanimate and abstract CNP subjects as being able to **do** things. Hyland (2006) argues that disciplines in the humanities and social sciences "give greater importance to explicit interpretation" than the harder sciences and rely on "the strength of argument" for their claims to be accepted (p. 240). Verbs, apart from functioning to narrate facts and report expository processes, play a central role in constructing interpretation and argumentation, projecting the author's stance and attitudes towards them in the meantime. In this sense, student writers' limited lexical variation of predicate vocabulary used with CNP

subjects also reflects their inadequate involvement in the meaning and knowledge making process. In other words, student writers need to be further apprenticed into the broader academic discipline and the research domains they write about in order to more effectively turn experiences into linguistic meaning. A more varied range of predicate vocabulary is also conducive to opening up greater possibilities for creative and effective meaning making and knowledge generation.

5.2.5.3. Lexical Sophistication

Based on the frequency range of the lexical items used to predicate CNP subjects obtained from COCA-Academic (see Section 3.5.2), this study sets up three frequency cut-off points as criteria for determining lexical sophistication: 1-50, 51-100, and 101 or more (frequency per million words); items with a frequency of 100 or fewer per million words are regarded as low-frequency.

TABLE 5.12

Frequency Distribution of Predicate vocabulary Taken by CNP Subjects

		MDC	**PRAC**
	Frequency cut-off	Percentage	Percentage
understanding	1-50	27%	52%
	51-100	n/a	5%
	≥101	73%	43%
analysis	1-50	12%	34%
	51-100	12%	7%
	≥101	76%	59%
lack	1-50	20%	34%
	51-100	13%	14%
	≥101	67%	52%

Table 5.12 presents the frequency distribution of the predicate vocabulary used with CNP subjects expanded from each of the three head nouns. The overall pattern of contrast that can be observed is that the lexical items used with CNP subjects feature a higher representation of low-frequency items in the expert corpus than in the student corpus for CNPs with each head noun: 57% vs. 27% for *understanding*, 41% vs. 24% for *analysis*, and 48% vs. 33% for *lack*. It can also be seen that the major difference lies in the proportion of items with a frequency of 50 or fewer instances per million words: 52% vs. 27% for *understanding*, 34% vs. 12% for *analysis*, and 34% vs. 20% for *lack*. Therefore, this contrast in the amount of low-frequency lexical items used with CNP subjects across the two corpora serves to mark the different degrees of lexical sophistication: the lexical items used with CNP subjects in the student corpus exhibits weaker lexical sophistication than those in the expert corpus.

It is usually the case that low-frequency words entail more delicate and detailed semantics than their high-frequency counterparts from the same semantic field. Take the whole array of reporting verbs for example, which include such words as *say*, *tell*, *speak*, *talk*, *inform*, *report*, *notify*, *discuss*, *verbalise*, *pronounce*, and the list could still go further. Among these reporting verbs, *inform*, *report*, *notify*, *discuss*, *verbalise*, and *pronounce* occur with lower frequencies than *say*, *tell*, *speak*, and *talk*, which are usually defined via the higher-frequency prototypes but with specific semantic, pragmatic, and situational qualifications added to them. For instance, according to the online *Longman Dictionary of Contemporary English*, the core sense of ***discuss*** is "to **talk** about something with another person or a group in order to exchange ideas or decide something" (http://www.ldoceonline.com/dictionary/discuss).

Thus, at least one of the reasons for the inclusion of low-frequency verbs in academic discourse relates to their greater semantic delicateness. The processes of

narration, description, observation, and argumentation via the use of lower-frequency (i.e. more sophisticated) verbs, are made detailed and accurate in meaning. Meanwhile, the already highly grammatically complex nominal subjects are made semantically sophisticated through low-frequency predicate vocabulary. It is conceivable that assigning varied and expanded meanings to inanimate and abstract nominal subjects would be cognitively challenging and novice student writers are still on their way to becoming more sophisticated in substantiating this type of subject-predicate clause structure.

Again, Hinkel's (2003) characterisation of using private verbs and public verbs (terms referring to the semantic categories of most reporting verbs) as representing immature L2 writing needs to be called into question. Her study showed that college-level EFL writers used significantly more private verbs (i.e. mental verbs like *think*) and public verbs (i.e. communicative verbs like *say*) than native speakers, stating that "public, private, and expecting/tentative verbs are very common and lexically simple" (p. 293). However, the present study has found it common for these verbs to be used with abstract CNP subjects in the expert corpus, including *indicate*, *suggest*, *refute*, and *accentuate*. In fact, the use of mental and communicative reporting verbs has been well recognised as an important lexical category in academic writing (Thompson & Yiyun, 1991), serving to present not only factual information like data and results but also the author's judgement and evaluation that is inherent in the verbs selected. Employing the same argument for the legitimacy of using copular verbs in academic writing, it can be said that the problems with EFL writers' use of verbs as those identified by Hinkel are not so much concerned with which semantic category of verbs to use as with how semantically elaborated and precise words are selected. Developing abstract processes

and entities into value-added speech and thought requires the use of lexis loaded with more delicate, hence more sophisticated semantics.

5.3. Summary of Chapter

This chapter has presented the results of five aspects of noun phrase complexity in student writing, as compared with expert writing: 1) the tendency of nouns to be postmodified, 2) the use of postmodifiers of different structural types, 3) multiple postmodification in the complex noun phrase, 4) complex noun phrases serving as clause subjects, and 5) use of predicate vocabulary for CNP subjects. Correspondingly, the following findings have been obtained.

First, it has been found that student writers tend to use postmodification in producing complex noun phrases in a way that is similar to expert writers, with an average of 42% of noun phrases having at least one postmodifier. Meanwhile, the six head nouns have been found to show different tendencies to take postmodification, which are also similarly represented in the two corpora.

Second, although both student and expert writers use preposition phrases as the major postmodifier type, student writing shows a slightly higher proportion of clausal postmodifiers, which is likely to be caused by students' strong preference to use clausal postmodifiers for the head noun *information*. The examination of using prepositional postmodifiers headed by *of* and other prepositions has revealed that students use a higher proportion of prepositional postmodifiers headed by *of* and a lower proportion of those headed by other prepositions than expert writers. The discovery that students sometimes select prepositions that are semantically incompatible with certain head nouns also needs to be taken into account when explaining students' preferences for prepositional postmodifiers. In the choice between finite and non-finite relative clauses

as clausal postmodifiers, student writers are found to prefer the former to the latter compared with expert writers, although this preference is not consistently found among all head nouns. This preference is in part explainable by the finding of students' frequent use of relative clauses introduced by *which*.

Third, the examination of students' use of multiple postmodification in complex noun phrases has found that student writers produce shorter and fewer levels of consecutive postmodification than expert writers. However, multiple postmodification in expert writing is found to centre around two consecutive postmodifiers, demystifying the perception that multiple postmodification is usually extremely complex. The textual significance of using complex noun phrases with multiple postmodification is that an underdeveloped noun phrase complexity has been found to contribute to a motion of discourse that lacks coherence, logical connection among ideas, and authorial evaluation of the themes under investigation.

Fourth, the examination of complex noun phrases serving as clause subjects has revealed that student writers use fewer CNP subjects than expert writers in general, with CNPs headed by *analysis* showing the greatest gap between corpora. However, the two corpora exhibit a significant difference in the length of subject CNPs headed by all three head nouns. A further test of greater-length CNP subjects shows that there are far more occurrences of longer subjects with all three head nouns in expert writing than in student writing. A brief discourse analysis was performed of students' use of markedly short and unelaborated CNP subjects from the perspective of information structuring. It is argued that students' apparent lack of syntactic sophistication to use CNP subjects to pack information either for introducing new elements into the discourse or for summarising those previously mentioned could be considered as

implicitly grounded in their underdeveloped awareness of discourse-oriented principles of information structuring in academic discourse.

Last, the transitivity patterns and lexical richness of the predicate vocabulary taken by CNP subjects were examined. Results show that student writers resemble expert writers in the transitivity patterns of the lexical resources predicating CNP subjects, with transitive verbs being the most represented category, followed by copular verbs and intransitive verbs. However, student writers are found to use a smaller variety of transitive verbs than expert writers, as well as fewer instances of the copula + adjective + (optional) complement construction. Students' weakness in deploying lexical resources from this transitivity pattern has been discussed. In terms of lexical richness of the predicate vocabulary, student writers use a smaller range of vocabulary to predicate all three sets of CNP subjects and the lexical items they use show a weaker degree of lexical sophistication.

Now that it has been seen noun phrase complexity is significantly insufficient in student writing, it would be reasonable to presume that student writers may rely heavily on clausal structures which are more typical of speech as the major discourse developing strategy in light of the developmental prediction (Byrnes, Maxim, & Norris, 2010; Ortega, 2003). However, considering the fact that clauses are used differently in speech and writing and that creating complex meaning relations such as causality and concession is crucial for scholarly argument building, it is still uncertain whether students use fewer clausal structures and how they use them. The following chapter will report and discuss students' use of various types of adverbial clauses and explore its discourse-related underpinnings. Hopefully, the analyses could provide some explanatory power for the results of syntactic complexity measures and frequency counts.

CHAPTER SIX

CLAUSE COMBINATION IN STUDENT WRITING

6.1. Introduction

Chapter 4 (Section 4.2) has demonstrated that student writing shows significantly lower scores in most measures of subordination. This chapter presents an in-depth textual examination of clause combination to better understand this underperformance. As described in 3.5.3, the examination of clause combination in this study focuses on the use of circumstantial adverbial clauses, both finite and non-finite. However, the findings obtained through the subordination-based measures included in the L2 Syntactic Complexity Analyzer (L2SCA) could be taken as a solid basis, because these measures do not differentiate among distinct types of dependent clauses. Moreover, only finite adverbial clauses are included in the subordination measures, leaving non-finite structures unaccounted for with respect to clause-combining features in student writing. Therefore, the study makes the case for a more delicate quantitative examination of clause combination as the basis on which to perform textual analysis.

This chapter presents frequency-based comparisons of adverbial clauses of different structural and semantic categories between the two corpora and then provides textual evidence and contextual explanations for the quantitatively displayed patterns. Given the significantly lower scores of the overall measures of subordination in the student corpus (see Table 4.3), the frequency-based comparisons were intended to see if student writers also use fewer adverbial clauses of different structural and semantic categories than expert writers. Meaningful textual and contextual analyses could then be based on prominent results of the comparisons. The qualitative textual and contextual analyses aim to explain the differences (if any) from the perspectives of

discourse semantics, information management, speech-writing interface, and cross-rhetorical transfer, considering their potential impacts on syntactic features. Attempts will also be made to provide pedagogical implications where clause combination can be enhanced and improved for clearer and more effective meaning making.

6.2. Results and Discussions

This section reports the results of the frequency comparisons of both finite and non-finite participle clauses used in the two corpora before providing detailed textual analyses of the two clause types used by student writers. Discussion and interpretation of the findings will be carried out alongside the quantitative and textual evidence, starting with finite adverbial clauses and then proceeding to participle clauses.

6.2.1. Results and Discussion for Finite Adverbial Clauses

This section presents the normalised frequencies (occurrences per million words) of finite adverbial clauses of three semantic categories (i.e. reason, concession, and condition) in the two corpora and results of statistical comparisons along with discussions of the results (see Section 3.5.3.2.1 for analytical details). The next three sections present discourse-related analyses of students' use of this clause type. The last section is a summary of this part.

6.2.1.1. Frequency Comparisons

As described in 3.5.3, the frequencies of three subordinators (or conjunctions) respectively representing the three semantic categories of finite dependent clauses were counted and compared between the two corpora. The three subordinators selected for

analysis are *because*, *although*, and *if* as the canonical subordinators for causal, concessive, and conditional clauses respectively.

As can be seen from Table 6.1, with the three semantic categories taken together, there are significantly fewer subordinators in the student corpus than in the expert corpus (a gap of 540 subordinators per million words in total), indicating that student writers produce significantly fewer finite adverbial clauses to explicate the logical and semantic relationships in presenting argument. This finding is consistent with the results of the subordination measures in the L2SCA which also showed significantly lower scores in student writing (see 4.2).

TABLE 6.1

Frequency Distribution of Representative Subordinators across Corpora

Semantic category	Representative subordinator	PRAC	MDC	*Log-likelihood*
Reason	*because*	1,100	718	80.868***
Concession	*although*	867	443	139.736***
Condition	*if*	874	1,140	-35.235***
Total		2,841	2,301	56.814***

Note. Frequency was normalised to instances per million words.

***Difference is significant at the $p < .001$ level.

However, despite overall under-representation of adverbial clauses in the student corpus, a closer examination of the distribution pattern for the three semantic categories reveals interesting variation: while both causal clauses exemplified by *because* and concessive clauses exemplified by *although* are significantly underused in the MDC, students' use of conditional clauses exemplified by *if* shows a significant

over-representation. This variation between students' use of causal and concessive clauses and that of conditional clauses may be explainable in terms of the varying degrees of informality of the three circumstantial adverbials. According to Biber et al. (1999, p. 842), among the three subordinators' distribution in informal conversation, adverbial clauses led by *if* have the highest frequency (3,400 instances per million words), followed by *because* and *although* clauses (respectively 2,600 and 200 instances per millions words). Therefore, the supreme prominence of *if*-clauses in informal registers marks its highly situated nature and relative effortlessness of reasoning, thus prompting students to constantly employ them in their ongoing composition. Besides, unlike causal and concessive clauses which mainly present evidence-based, content-oriented argumentation, conditional *if*-clauses have been found to perform a range of interpersonal discourse functions in academic writing (Warchał, 2010). It might be these interpersonal aspects of *if*-clauses—closely associated with involved registers such as conversation—that is also accountable for students' overuse of this semantic category.

By contrast, student writing shows significant underuse of *because*- and *although*-clauses representing the semantic relationship of reason and concession respectively. These two categories of semantic relations constitute important means to construct coherent argument structures. Among the two, the frequency gap in the use of *although*-clauses is even larger than that for *because*-clauses, with expert writers using *although*-clauses almost twice as frequently as their student counterparts, which indicates that the concessive relation poses the greatest challenge for student writers. This finding testifies to the argument advanced by Kortmann (1991) that the concessive is "the most complex of all semantic relations that may hold between parts of a discourse" (p. 161).

The prominence of causal and concessive relations in published academic writing thus lends empirical legitimacy to measures of clausal complexity against the rising criticisms levelled at these measures on grounds of the oral nature of clause combination. As discussed in 2.4.2.4, the appropriateness of using certain grammatical features to gauge writing complexity cannot be solely based on frequency differences between speech and writing, given the radically different semantic force and discourse functions of adverbial subordinators in the two modes of discourse (Schleppegrell, 1996).

In summary, the break-down analysis of adverbial clauses has shown that although student writers underuse argument-building adverbial clauses overall, which is consistent with the results obtained from the subordination measures, the real challenges contributing to students' underuse mainly come from the more epistemologically and cognitively demanding semantic categories of causal and concessive clauses. Thus, students' subordination-based syntactic complexity needs to be explained in light of the extent to which they are able to articulate and complexify argumentation in their writing. In what follows, the attested quantitative weaknesses in students' use of causal and concessive clauses is related to the use and complexity of argument advancement in order to explain how surface-level syntactic complexity in the use of adverbial clauses reflects and is a result of higher-order epistemological, cognitive, and disciplinary attainment.

6.2.1.2. Weakness in Causal Exposition

The underrepresentation of *because*-clauses in student writing suggests that student writers do not provide adequate grounds and justifications for the claims or courses of action presented in their dissertations. It might also be possible that they

have left much underlying causality unarticulated at the syntactic level, only deducible by the reader through observing the potential relationship "out there" beyond the syntactic surface. Research into the discourse situations in which *because*-clauses are used has found that in the more formal registers of written texts they "regularly emerge after rhetorical relations of contrast and negation, or more generally, after propositions that are distinct from potentially shared expectations….[as well as] after strong evaluations or after other claims that are presented as remarkable or stronger than normal" (Ford, 1994, pp. 531-548). Such discourse environments in which *because*-clauses are used can be easily attested in the expert corpus, although they appear both before and after the matrix clause, such as the following three examples in conjunction with negation, claim, and evaluation respectively:

(1) (Text 27, PRAC)

Account→ *Because the participants who participated in this experiment were all adults,*

Negation→ *we could safely assume that they had no difficulty in making relative size judgments on daily objects.*

In (1) the author anticipates the possibility that "making relative size judgements on daily objects" could be a challenging task for the cognitively underdeveloped but counters the possibility by explaining that the participants were adults who would not face such a challenge.

(2) (Text 112, PRAC)

Account→ *Because words with multiple meanings constitute a large part of L2 vocabulary,*

Claim→ *ESL teachers should find ways to help accelerate learners' access to the correct meanings of a polysemous word used in different contexts.*

In (2), before making the claim about teachers' obligatory pedagogical practice in vocabulary instruction, the author provides a factual account of the characteristics of L2 vocabulary to justify the claim to be immediately advanced.

The short matrix clause in (3) is highly evaluative of the author's own finding by using the strong judgemental modifier *important*, which necessitates an equally strong and cogent justification, hence the lengthy *because*-clause detailing the reason why the finding is an important one. Indeed, circumstances abound in the expert corpus where evaluations and claims are supported by justification and explanation, as professional writers are always prepared to defend their claims and arguments by foreseeing potential suspicion and misunderstanding from the readership.

(3) (Text 21, PRAC)

Evaluation→ *This is an important finding,*

Account→ *because it allows us to better understand one reason why teachers have negative attitudes to much published (and, in their eyes, overly theoretical) research and why, for several teachers in this study, reading research was an activity they engaged in only when they had to (e.g., for promotion).*

However, as will be exemplified later, causal relationships are not always explicitly addressed in student writing even where the potentiality of establishing causal links between statements can be sensibly deduced by the reader. In addition, even though student writers do employ *because*-clauses to a great degree in their dissertations, as will be shown, the complexity of the logical link between the *because*-

clause and the matrix clause is rather weak, largely due to the lack of within-clause noun phrase complexity constituting the major ideational force of argument. In what followsm, texts from the student corpus will be extracted for examining circumstances in which student writers fail to syntactically present explicit causality where it would be necessary to do so.

Consistent with the finding that sentences are significantly shorter in the student corpus (see Table 4.3), it is easy to find short sentences that are chained one after another without enough syntactic and textual signals indicating the semantic relationship in between. The following excerpt gives a flavour of the grammar-borne weakness of logical linkage in student writing, with reference to the necessity of providing grounds for claims and negation:

> (4) [1]*Most of the time, we make use of the reading material through processing the input.* [2]*Answers cannot always be obtained directly from the text.* [3]*This requires the test takers to have the abilities to induce and deduct the input.* [4]*However, the reading tasks of the four TIE-1 past papers do not reflect the candidates' inductive skills.* [5]*Thus, the reading component fails to reflect their overall reading abilities.* (Text 02, MDC)

The general syntactic style of this extract is characterised by short sentences: the average sentence length is 14 words, with the longest being 18 and the shortest being 8. A perusal of this extract seems to suggest two large communicative goals: to justify the cognitive demand of reading comprehension for test takers and to point out the shortcomings of a particular reading test. The student writer employs two to three short sentences to develop each goal: the first goal with the nucleus in Sentence 3 is based on a claim expressed in Sentences 1 and 2, and the second goal with the nucleus in Sentence 5 is based on the claimed fact expressed in Sentence 4.

However, these causal relationships are only hidden between the lines, without being made explicit by using clause combining strategies to link together the propositions contributing to the local coherence associated with each goal (Zhang, 2014). Local coherence is thus broken by using two independent sentences, imposing unnecessary mental breaks in the reader's comprehension. Analysing the first two sentences which form the claim for advancing the argument in Sentence 3, it could be inferred from what the student writes that the negative claim that "*answers cannot always be obtained directly from the text*" seems to be predicated on the foregoing claimed fact that "*most of the time, we make use of the reading material through processing the input*". However, what she has failed to do in attempting to convey this plausible causality is reflected in the fact that there is not adequate signalling for the link to be made explicit. That is, the couple of short sentences contributing to a single aim fails to be combined into a larger unit by clause combining strategies such as using conjunctions (e.g. *because*) or non-finite clauses.

Besides, the reader needs to take constant pauses at short intervals imposed by the periods, with each punctuation unit (Chafe, 1988) containing only partial information of a more complete argument structure. Experienced readers who are accustomed to extended and complex information when doing subvocal or speed reading would find these constant pauses unexpected and somewhat detrimental to processing the logic flow of the discourse. The lack of syntactic marking of the causality between the propositions is likely to result in monotonous reading and slow down comprehension with the reader having to constantly guess at the potential causal links hidden among the short statements.

Without a major renovation of the student writer's own wording, Ex. (4) could be restructured with stronger causal linkage to present the two goals:

(4-Restructured) *Because/since we make use of the reading material through processing the input most of the time and answers cannot always be obtained directly from the text, this requires the test takers to have the abilities to induce and deduct the input. However, as the reading tasks of the four TIE-1 past papers do not reflect the candidates' inductive skills, the reading component fails to reflect their overall reading abilities.*

Although it may seem somewhat far-fetched to impose causal interpretation on the original text, the rewritten version at least helps smooth information flow and reduce logical breaks caused by frequent full stop punctuation.

Sometimes the semantic relations between short sentences as in Ex. 4 can be made explicit by conjunctive adverbs (or disjuncts) instead of using clause combining devices. In expressing the causal relationship between Sentences 4 and 5 in Ex. 4, the student writer uses the conjunctive adverb *thus* to mark the link. Conjunctive adverbs such as *thus, however,* and *therefore* are usually used to introduce a different theme than the previous chunk of discourse, and global (as opposed to local) coherence can be established between themes. Here the student writer was employing a conjunctive adverb that is usually used to develop global coherence to connect two sentences that are locally coherent—in this case the two sentences contribute to a single aim of argumentation. With the two sentences separated, it would be difficult for the reader to immediately realise that not only Sentence 4 but the combining force of Sentences 4 and 5 stand in an adversative relation with the previous argument unit. Therefore, it is clause combining devices that should have been used to overtly produce the causal link between Sentences 4 and 5 to establish local coherence.

It also needs to be noted that the two claims in Sentences 2 and 5 are both made with negation (*cannot* in 2 and *fail* in 5), necessitating immediate justification and

explanation for the negated claims, which can be best realised by welding supportive statements through clause combining devices (Ford, 1993). However, the student writer fails to provide this immediate account in the locus of the negative claim clause, leaving isolated a statement that could have been merged into the coherence scope of the claim through a causal clause to establish a claim → account rhetorical pattern. Further examination of the student corpus for other rhetorical patterns in which a causal clause usually occurs has revealed cases in which a rhetorical move such as strong evaluation stands alone without enough account provision. Consider the following extract from the literature review chapter of an MA dissertation.

> (5) *English Tense is an important and also a controversial grammatical category in Modern English grammar. In English, tense originates from the Latin word Tempus with the meaning of time. The first person who confirmed the concept of tense is the philosopher Protagoras in the 5th century BC (Liu Ruiqing 2002:12). Many contemporary English grammarians define tense a grammatical category to indicate time distinctions.* (Text 03, MDC)

Here the student writer evaluates "English Tense" as an "important" and "controversial" grammatical category. It would be natural for the reader to expect further elaboration of the extent to which this grammatical category is "important" and "controversial" given the strong evaluation made about it. However, no such evidence is provided either within the statement or in more extended context, thus displaying a sense of arbitrariness and ungroundedness of the evaluation made. This could be due to student writers' lack of pressure to constantly provide supportive grounds for claims and evaluation or due to their underdeveloped ability to find adequate evidence to support their argument. Although providing supportive grounds for evaluative claims is not just limited within clause combinations and can extend to the wider discourse, the underuse

of *because*-clauses in the student corpus does point to an underrepresentation of using reason adverbial clauses as a means to fulfil the local claim → account argument move.

While Biber et al. (2011) have argued against using linguistic features that are more common in conversation—of which *because*-clauses are a typical example—to measure syntactic complexity of academic writing, the finding of this study that student writers tend to use significantly fewer causal clauses than expert writers serves to present the issue for careful re-examination. Biber et al. seem to have overlooked one important aspect of the difference in the use of *because*-clauses between conversation and academic writing, that is, the "semantic force" of the causal relationships encoded between the matrix clause and adverbial clause in informal speech is different from that which is often found in academic writing (Schleppegrell, 1996, p. 272). Schleppegrell points out that *because*-clauses in spoken discourse "have a range of functions related to interactional concerns…mak[ing] pragmatic contributions to the ongoing interaction and structuring of discourse, and their semantic contributions are less in focus".

By contrast, conjunctions in academic writing "are more readily interpreted as markers of meaning relationships in texts while the pragmatically oriented functions "rarely occur in academic writing" (ibid, pp. 272-273). Simply put, semantic relationships marked by the conjunctions in academic writing necessitate higher-ordercritical thinking as compared with everyday conversation. From the perspective of reasoning logic, transferring oral-style *because*-clauses into academic writing may result in arguments with informal fallacy (Toulmin, 1958), given the vulnerability of everyday logic to speciousness.

For example, providing personal experiences and anecdotal examples in the *because*-clause, while common in everyday meaning negotiation, cannot be regarded as forming logically forceful causal relationships with a claim, such as the following

sentence written by an ESL college student cited from Schleppegrell (1996, p. 276), where the student writer gives an example of "flexibility" rather than providing a justification for it:

(6) *Schedules [in American schools] are flexible because students who don't like history can take geography instead*.

Thus, the tendency of student writers to underuse *because*-clauses could be interpreted in terms of their conscious restraint from transferring oral characteristics of adverbial clauses to academic writing. In this sense, the underuse can be partly understood as students' avoidance to provide semantically forceful causal reasoning considering its requirement of the writer's epistemological and disciplinary development.

The above analysis of student writing samples taken from the MDC has provided important revelations about students' underuse of causal adverbial clauses exemplified by *because* from the perspective of discourse semantics and argument building. First, student writers evidently fail to make explicit the potential causal links hidden among short sentences that are characteristic of overall discourse (consider the significantly shorter mean length of sentence); these links could and should have been syntactically marked using clause-combining strategies. Second, sometimes strong evaluations and claims in student writing are not provided with adequate supportive grounds, resulting in weakness in both syntactic and argument complexity. Third, the epistemologically and cognitively challenging demand on producing logically sound and forceful causal links between propositions inherent in convincing argumentation may have daunted novice student writers in their attempts at employing *because*-clauses.

6.2.1.3. Weakness in Concessive Clauses

The concessive *although*-clause is the other semantic category of adverbial clauses (among condition, reason, and concession) where the student corpus shows a significant underuse. Different from clauses led by *because* and *if*, which are more commonly used in conversation than in academic writing, *although*-clauses together with concessive *while*-clauses are mainly used in academic prose (Biber et al., 1999, p. 842), indicating their strong association with this register. Student writers' underuse of *although*-clauses also points to an important area of argument presentation in academic writing where they significantly lag behind expert writers: the recognition and acknowledgement of the inadequacies of certain methodological procedures or potential criticisms and disagreement from the readership about the author's line of argument, and equally importantly the rebuttal of these threats and accusations.

This oppositional rhetorical step, called "Countered-Rebuttal" in Crammond's (1997, 1998) model of the semantic networks of argument development based on the general structure of "claim ↔ support" (Toulmin, 1958), is among the elaborative rhetorical structures characterising argumentative texts. Countered-Rebuttal requires the presentation of both the potential threats to one's claim and their refutation in defence of the claim. While this can be developed through a number of independent sentences stating the different elements required in presenting Countered-Rebuttal, employing clause combination enhanced by the concessive/adversative relation would produce more coherent and compact argumentation development. Therefore, analysing the use of *although*-clauses in the student corpus, the most representative and frequent form of the concessive relation, can provide direct insights into the extent to which student writers foresee potential challenges from the extended readership and defend their claims and research actions in due course.

It needs to be noted that using concession for realising Countered-Rebuttal is prototypically operative in fundamentally argumentative essays. But in academic writing, concession is also used for presenting contrastive situations other than rebuttals and it can appear in any rhetorical section of the research article or dissertation (e.g. reporting contradictory findings in the "Results and Discussions" section). Epistemologically, concession serves two functions: 1) the *although*-clause displays the author's knowledge of the existing inadequacies and alternative viewpoints concerning certain aspect of the topic in question and how they can be detrimental to the author's own line of argument or methodological decision; and 2) the matrix clause shows the author's judgement that the negative influences can be mitigated and even excluded by demonstrating their insignificant impact for the study at hand. At the same time, demonstrating such knowledge by way of constructing the concessive relationship fulfils the discourse semantic demand of providing Rebuttal and Countered-Rebuttal essential to high-quality argumentative elements (Crammond, 1998). Thus, student writers' underuse of the concessive *although*-clause can be partly explained by their insufficient disciplinary attainment necessary for identifying competing facts and ideas for constructing a contrastive stance.

Similar to what contributes to student writers' underuse of *because*-clauses as illustrated above, their inadequate use of *although*-clauses is also largely interpretable in discourse-related terms, particularly in association with students' disciplinary and epistemic accomplishment and their sensitivity to the cognitive demands of academic discourse semantics. Being able to foresee possible challenges from the readership first requires writers to recognise the weaknesses and limitations of a particular theory, method, or finding—either that of their own or of others—in the context of the discipline and research domain under examination. Since students' insufficient use of

although-clauses is closely linked to such contextual factors as underdeveloped disciplinary knowledge and apprenticeship into the research area, it is difficult to provide clear textual evidence to explain the underuse pattern. Where exactly an *although*-clause (or other concessive resources) is needed may not be objectively decided and is only up to the author's own observation on the semantic necessity to incorporate contradicting elements in the unfolding text. In addition, much of the ability to recognise potential audience challenges and contradictories is dependent on the writer's critical thinking capacity. This corresponds to Kortmann's (1991) comment that the concessive is the most complex of all semantic categories of discourse semantics.

It now becomes clear that what is included under the rubric of subordination-based syntactic complexity should necessarily be considered from the angle of conceptual fulfilment of certain discourse semantic elements required of high-quality argumentation. The use of adverbial clauses is not a "complex" process in itself, but identifying complicated logical and semantic relations in the textual and contextual surroundings and presenting them in explicit language is where "complexity" comes in.

6.2.1.4. Semantic Force Enhanced by Noun Phrase Complexity

Another important factor that may have dampened students' enthusiasm to produce *because*- and *although*-clauses could be due to the clause-internal noun phrase complexity both in the adverbial clause and the matrix clause. Consider the following example from the expert corpus:

> (7) *Although explicit knowledge of grammatical rules does not guarantee corresponding changes to implicit grammatical accuracy, research since the mid-1980s has suggested an indirect, but facilitative, role for explicit*

knowledge in improving the rate and ultimate level of L2 attainment (Lightbown and Spada 2006). (Text 01, PRAC)

As can be seen, both the *although*-clause and the matrix clause of this sentence contain extended CNPs (underlined) in the subject and object positions and the semantic relations between the CNPs in each clause hinge on the lexical verbs (i.e. *guarantee* and *suggest*). As such, it would already have been highly challenging for the reader to understand the meaning at the clause level. However, the complexity is further intensified by the concessive relationship between the two clauses, making the ideational meaning even more difficult to decode. It would not be difficult to agree that the psycholinguistic process of understanding this kind of complex sentence (as called in traditional grammar) is a highly complicated one. In the case of the above example where the *although*-clause is in sentence-initial position, readers have to readily keep in mind, while reading the concessive clause, that there is a competing or contradictory relationship between what is being read and what is going to appear in the matrix clause. Therefore, the reader has to come to terms with the two types of complexity simultaneously: strenuous decoding of the ideational meaning in the complex noun phrases together with the semantic relationship between the CNPs (realised by the lexical verbs) and between the clauses (realised by the use of *although*). Thus, the reader's cognitive energy is constantly at full blast in dealing with these two types of complexity.

A similar level of cognitive investment is necessary for the production of clause combinations featuring the kind of syntactic and semantic complexity shown above. This requires the writer to be highly resourceful and skilful in packing and abstracting experiential processes into complex noun phrases and to be highly observant in

identifying the underlying logical relationships between the meaning made at various levels.

Not only have student writers used fewer causal and concessive clauses, the complex sentences they produce usually display a marked weakness in NP complexity and logical force. Typical expression of reason and concession in student writing is exemplified as below, with complex noun phrases underlined:

(8) *Because oral class is new thing for most students, the researcher has an interview with <u>15 students selected from the experimental group</u>.* (Text 39, MDC)

(9) *Because the L2 now appears 'set in stone', <u>the term fossilization</u> was used to describe this point.* (Text 22, MDC)

(10) *Although he had produced a descriptive draft according to the requirement, the piece was short in length and poor in content.* (Text 01, MDC)

(11) *Although there are <u>some previous studies on the simple present tense in prototype theory</u>, it is not comprehensive and deep enough.* (Text 03, MDC)

The underline visualisation shows that these four sentences, each containing an initial-position adverbial clause, are characterised by scarce use of CNPs in either the adverbial or matrix clause, and therefore weak semantic force is formed between the two clauses. In some cases, in order to convey clearer and more concise messages, alternative expressions without using an adverbial clause might have been better than clause combination Thus, dispensing with the *although*-clause, Ex. 11, which does not seem to entail a strong concessive or contrastive relation between the adverbial clause and the matrix clause, could be more directly expressed as a "simple sentence" through information packing, with CNPs underlined:

(12) *Previous research on the simple present tense in prototype theory is not comprehensive and deep enough.*

The revised version retains the major focus of message which the student's original clause combination is intended for, i.e. something to the effect that previous research is not good enough. The point being made here is that the logical link encoded in clause combinations in student writing can sometimes be so weak that using an independent clause with skilful information packing is enough to get the meaning across. In academic writing the causal and concessive relations are more frequently reserved for combining clauses involving more complicated participants and processes. The level of ideational and logical sophistication in student-produced clause combinations is limited by both their assessment of the research domain being written about and their awareness of the prominence of noun phrase complexity at the clause level.

6.2.1.5. Summary of Finite Adverbial Clause Combination

Interested in seeing how the use of specific dependent clause types differs between students and expert writers, this section has unpacked the subordination measures included in the L2SCA to examine the different frequencies and textual features of an important component of these measures, the circumstantial adverbial clause. Frequency results show that among the three semantic categories of condition, reason, and concession (respectively represented by three canonical subordinators *if*, *because*, and *although*), student writers use significantly more conditional clauses than expert writers but their use of the other two categories is significantly lower. Both textual and contextual factors have been invoked to account for students' underuse of the causal and concessive relations. It has been found that student writers sometimes fail to employ clause combining devices to make explicit the causal relationships left

hidden among short statements lacking explicit textual or logical links. More crucially, it has been shown that students are not always prepared to provide immediate justifications for their claims and evaluations by using causal clauses and that their writing lacks considerate recognition of potential audience challenges and corresponding pre-emptive rebuttal of them through the use of concessive clauses. In addition, not only do student writers produce much fewer causal and concessive relations, observation of their actual production of *because*- and *although*-clauses has shown weak noun phrase complexity, which further undermines the overall logical nexus between the adverbial clause and the matrix clause.

Examining the performances of adverbial clauses of different semantic relations has contributed to a fine-grained understanding of the differentiations in the use of dependent structures, highlighting the necessity of accounting for the differences in syntactic complexity from the perspectives of such textual factors as argument presentation, register awareness, and text development as well as such contextual factors as disciplinary involvement (Li, Hyland, & Hu, 2017) and critical thinking (Stapleton, 2001). These perspectives signify a promising orientation for researching syntactic complexity that is more grounded in functionally and semantically based factors that prompt or impede its growth in student writing (Ortega, 2015).

The findings also have implications for the importance given to adverbial clauses as valid indicators of syntactic complexity for advanced academic writing. Although adverbial subordination has often been characterised as features of informal conversation and premature L2 writing rather than professional academic writing in frequency-based terms, the fact that its usage and functions in the two registers are radically different has been somewhat ignored. As Schleppegrell (1996) points out, subordinators in spontaneous speech are mainly interpersonal and pragmatic markers

whereas they bring ideationally complex statements into semantically forceful relations in academic writing. The role of subordination in academic writing should not be tarred with the same brush as its highly interpersonal and somewhat verbose features in conversation. Although subordination does not enhance ideational complexity at the clause level, the level of cognitive investment and subject knowledge required for decoding the complex relations embedded in the clause combination is substantial.

6.2.2. Results and Discussion for Participle Adverbial Clause Combination

Similar to the presentation of finite adverbial clauses, the frequencies of various types of participle adverbial clauses in the two corpora and their statistical comparisons are presented along with discussion of the quantitative results. (Refer to Section 3.5.3.2.2 for details of analysis.) Subsequently, discourse analysis of students' use of the clause type is carried out. This section concludes with a summary of the quantitative and textual findings.

6.2.2.1. Frequency Comparisons

As described in 3.5.3, the frequencies of present and past participle clauses in two sentence-referenced positions were counted and subjected to statistical comparisons. Table 6.2 presents the normalised frequencies of supplementive clauses and *with*-absolute clauses in the two corpora and the results of statistical comparisons between the two corpora.

TABLE 6.2

Frequency of Participle Adverbial Clauses across the Two Corpora

Position	Type of participle adverbial clauses	PRAC	MDC	*Log-likelihood*
Sentence-initial -*ing* clauses	supplementive	233	127	30.77***
	with-absolute	7	3	1.05
Sentence-middle and -final -*ing* clauses	supplementive	1,255	754	124.62***
	with-absolute	128	33	54.51***
Total -*ing* clauses		1,623	917	231.20***
Sentence-initial -*ed* clauses	supplementive	83	81	0.01
	with-absolute	5	15	-3.63
Sentence-middle & final -*ed* clauses	supplementive	354	227	27.51***
	with-absolute	46	29	3.33
Total -*ed* clauses		488	352	28.65***
Total participle clauses		2,112	1,270	209.78***

Note. Frequency was normalised to instances per million words.

***Difference is significant at the $p < .001$ level.

The comparisons have yielded a number of findings in students' use of participle adverbial clauses. It can be seen that expert writers used significantly more participle adverbial clauses than student writers for the majority of the comparisons, with only four comparisons not showing significant differences. Overall, 2,112 participle adverbial clauses per million words are used by expert writers while student writers use only 1,270 of them. Looked at separately, both -*ing* clauses and -*ed* clauses see a significantly stronger representation in the expert corpus than in the student corpus. However, there appears to be a wider gap between the two corpora in the use of -*ing* clauses than -*ed* clauses, as shown by the greater *log-likelihood* score for the former

(231.20 vs. 28.65). A closer look reveals that three of the four subtypes of *-ing* clauses have shown significant differences between the two corpora, with only sentence-initial absolute clauses introduced by *with* not being significant. In contrast, only one subtype of *-ed* clause has shown significant difference: sentence-middle and -final supplementive clauses. These findings indicate that *ing*-participle clauses present a much greater problem for student writers than do *-ed* clauses, a pattern that has also been observed by previous studies investigating EFL learners' use of participle clauses (e.g. Cosme, 2008; Granger, 1997).

In addition, of the four comparisons that do not show significant differences, three are concerned with the use of absolute clauses. In fact, as shown in Table 6.2, absolute clauses introduced by *with* in sentence-initial position are very infrequent in both corpora; only sentence-middle and -final *-ing* absolutes see a significant difference. Therefore, it can be summarised that *-ing* supplementive clauses in all syntactic positions, and *-ed* supplementive clauses in middle and final positions constitute the major areas where student writers lag far behind expert writers, thus deserving intense pedagogical attention. In contrast, absolute clauses introduced by *with*, except for middle and final *-ing* clauses, do not appear to pose practical challenges for student writers, probably due to their relative rarity in academic writing.

The finding that Chinese student writers underuse participle adverbial clauses, particularly *-ing* supplementive clauses in academic writing adds to the knowledge that these structures are challenging to novice EFL writers irrespective of their L1s. This is so because both students from East Asian L1 backgrounds (Hinkel, 2002, p. 57) and European languages (Cosme, 2008; Granger, 1997) have been found to underuse this non-finite clause type, pointing to its elusiveness caused by L2-specific and other non-crosslinguistic factors.

Four factors can be proposed to explain student writers' inadequate use of participle adverbial clauses: 1) underdeveloped awareness of their information-integrating and discourse-structuring functions, 2) inattentiveness to this clause type in academic writing, 3) fear of running the risk of dangling structure, and 4) domain-related underdeveloped ability to provide extra information using this clause type. While the first three factors are largely grammatical and generic, the last points to extra-linguistic, contextual constraints on the production of participle adverbial clauses.

First, the use of participle adverbial clauses, which makes up one of the means to reduce finite clauses to non-finite or phrasal structures (see Tannen, 1982, p. 39ff.), is able to integrate additional information closely attached to certain element in the matrix clause; the cohesion thus created seems natural and orderly. Virtually non-existent in informal conversation, these clauses are unique in written registers. Apart from enriching syntactic variety and contributing to a compact, integrated style, participle adverbial clauses are a stylish variant of their finite prototypes to combine clauses into a T-unit. Sentence-initial participle clauses are useful when writers need to present statements that do not bear highly explicit semantic relationships with what is in the matrix clause. When used in the medial position, i.e. after the subject of the matrix clause, they are a concise alternative to non-restrictive relative clauses. When appearing after the matrix clause, they are ideationally equivalent to a coordinate clause introduced by *and* but exhibit a closer semantic tie with the foregoing matrix clause. Although this clause type has already been taught in secondary schools in China, in most cases only its grammatical features are stressed (e.g. the use of present or past participle, the underlying logical subject, and the risk of dangling), its distinction from the use of finite clauses and discourse functions are rarely touched upon. Students who have not been given explicit teaching on discourse- and semantics-related features of

participle clauses would be unlikely to recognise their subtle functional properties, merely leaving the structures on their grammatical resource shelf as something of little value. Predictably, students would be more prone to resort to separate sentences or coordination rather than participle clauses where the latter could have been a better choice for the creation of more cohesive discourse flow. However, textual evidence is needed for this claim to be validated, which will be provided later in this chapter.

Second, it is possible that students intuitively perceive participle clauses to be uncharacteristic of academic writing, thus leading to insufficient emphasis placed on them. This is because participle clauses—while unique to the writing mode—are actually more frequently used in narrative genres such as fiction and news report than in academic writing (Kortmann, 1991, p. 2). Without explicit teaching, mere input from reading academic literature may not be able to draw students' awareness to this clause-combining feature (as with other important but easily neglected grammatical features), because it is not a feature that would cause great comprehension obstacle; meaning, rather than form, figures prominently in the reading process.

Third, students may have been concerned about the fact that supplementive clauses are notoriously tricky for being susceptible to dangling when the underlying psychological subject of the participle clause does not identify with the grammatical subject of the matrix clause. Although instances of unattached dangling structures are acceptable under certain circumstances (Biber et al., 1999, pp. 829-830; Kortmann, 1991, p. 9), students are normally warned against using them. This warning may have perpetuated a belief in students that attempting to use participle clauses may likely result in the frowned-upon dangling structure, which in the long run may have caused students to avoid them.

Finally, as with student writers' underuse of finite adverbial clauses reported above, immature ability to present and discuss domain-specific disciplinary knowledge and empirical findings should also be included to account for the underuse of participle adverbial clauses. Therefore, the underrepresentation of participle adverbial clauses may also indicate some aspects of students' still maturing disciplinary expertise and ability to put "unverbalised brainstorms" into "words for sober reflection, or discussion" (Chase, 1956, p. vii). To illustrate, the basic discourse functions of participle clauses used in initial and medial/final positions are to foreground and background the matrix clause respectively. Many cases of the final-position *-ing clauses* serve to provide highly detailed explication of what is said in the previous matrix clause, for example:

> (13) *The vast majority of research looks at academic bundles in English, and therefore this chapter focuses on English, discussing the emerging research which demonstrates the importance of this type of formulaic language in both academic speech and writing and the extent to which it varies in frequency, form, and function by mode, discipline, and genre.* (Text 36, PRAC)

In this example, the *ing*-supplementive clause comprises the vast majority of the whole sentence, detailing the manner in which "this chapter focuses on English". Being able to use a supplementive clause to expand on a general statement as such depends on the author's level of expertise and involvement in the complexity of the research domain being written about. Academic writing reflects not just the realistic complexity of what one knows, but of what one does in the research process as well.

To summarise, it can be seen that both linguistic and non-linguistic factors may contribute to student writers' underuse of participle adverbial clauses. While it could be difficult to bring about significant progress in students' disciplinary knowledge and involvement in the short run, pedagogical assistance should be able to raise their

awareness and understanding of the textual and functional circumstances to which participle adverbial clauses are applicable. The next section examines textual features of student writing to provide both empirical explanations for students' underuse of participle adverbial clauses and direct pedagogical implications for teaching this syntactic feature in EAP contexts.

6.2.2.2. Weakness in Participle Adverbial Clause Combination

While there may not be handy remedies for student writers not using participle adverbial clauses to present comment or further information on the matrix clause due to their lack of subject-matter expertise, underuse driven by students' incomplete awareness of the structure's textual usefulness can be better understood and improved through discourse analysis of student writing. It is hoped that strong textual evidence could be located where student writers miss out on possibilities for clause combination using participle adverbial clauses. Functional and contextual characteristics of the textual evidence will also be identified to pedagogically alert student writers to the possibility of using this clause-combining strategy.

As has been discussed, participle adverbial clauses are semantically attached to the matrix clause, but the relationship often remains implicit. Therefore, experienced writers would opt for this device when they intuitively perceive the potential relationship between propositions but do not see a real necessity to make it syntactically overt by using adverbial conjunctions. Besides, this clause type is especially pertinent when the adverbial message is not to be given much prominence in the information structure. However, since there has not been adequate description of the functional and contextual correlates for using participle clauses in place of their finite counterparts,

textual analysis of student writing can be only based on observation of the contextual circumstances where the device could have been employed.

Observation of student writing has found that the underuse is mainly attributed to unarticulated relationships between statements that syntactically stand alone from each other, only recoverable from careful reinterpretation of the surrounding context. While this observation pervades the entire text across different sub-genres of the dissertation (i.e. introduction, methodology, results & discussion), there seem to be specific syntactic and functional situations where the local discourse arrangement appears inharmonious and using participle clauses would considerably improve cohesion and coherence. Two contextual circumstances can be identified as correlated with students' non-use of supplementive clauses: providing background information (as in Ex. 14) and presenting multiple attributes relating to the same grammatical subject or subject matter (as in Ex. 15). In addition, absolute clauses (introduced by *with*) are usually neglected when they should have been used for connecting two segments conveying closely related ideas but with different grammatical subjects or themes (as in Ex. 16).

(14) *Over the past 30 years, leaner autonomy has drawn considerable attention of many scholars in the field of language teaching. Learner autonomy is now commonly accepted as the important content and the ultimate goal in college English teaching*. (Text 39, MDC)

(15) *Long-term memory is a permanent accumulation of one's filtered experience and knowledge. It holds both our declarative knowledge and procedural knowledge.* (Text 20, MDC)

(16) *For linguists and grammarians, the study of English tense has always been one of the main subjects. A large number of books and articles focus on the English tense in recent years*. (Text 03, MDC)

Again, a rough reading of the three examples gives an overall impression of weak syntactic complexity in terms of sentence length. A closer look reveals either repetitiveness and redundancy (Ex. 14 and 15) or broken logical linkage (Ex. 16) at the between-sentence discourse level. The information contained in each excerpt can be integrated into a more coherent single sentence by using participle adverbial clauses to encode background and other circumstantial information. Detailed textual and contextual analyses of the three excerpts are provided below.

In (14), the student author emphasised the importance of learner autonomy as an educational goal after introducing its historical significance in English language teaching research. In terms of the discourse semantics of the short excerpt, the historical background encoded in the first sentence does not necessarily lead to a natural interpretation of its importance in practical language pedagogy as conveyed in the second sentence. In fact, the logical relationship between these two propositions, if any, seems quite unclear: it could not be deduced just from the sequencing of the two sentences whether it is the widespread research coverage of learner autonomy during the 30 years that has led to its recognition as an important educational goal or the other way around. Therefore, it might be for this reason that the student writer ends up sequencing the two apparently unrelated propositions in two independent sentences, leaving the logical relationship between them untouched on. As a result, it would be difficult for readers to understand the point being emphasised by the author, not least because the two sentences start with the same grammatical subject (i.e. *learner autonomy*) which further blurs the intended focus. The student might not have realised

that the participle clause can be an effective strategy for incorporating background information of a particular subject matter that is to be immediately addressed in more detail. Thus, (14) could be restructured as:

> (14-Restructured) *Having drawn considerable attention of many scholars in the field of language teaching, learner autonomy is now commonly accepted as the important content and the ultimate goal in college English teaching*.

Thus, the supplementive clause not only encodes the historical information concerning learner autonomy but also effectively brings this information to foreground later predication of the concept, which can then stand as the matrix clause as well as the focus of the combined sentence. Although still remaining semantically implicit in relation to the matrix clause, the supplementive clause recedes into a secondary, subordinate position, making the matrix clause stand out in the information flow. At the same time, using the supplementive clause can mitigate against the awkward problem of repetitive subjects and enhance grammatical complexity as embodied in a greater structural variety.

Example (15) presents two attributes of "long-term memory" in two sentences, with the anaphoric *it* in the second sentence referring back to the grammatical subject of the previous sentence; therefore, the two sentences serve to characterise the same subject matter. It is possible that the student has read a great deal about long-term memory from the literature and was eager to display her knowledge and understanding of the concept through writing. But the problem confronting the student could be that she knows so much about the concept that she cannot help presenting as many aspects of the concept as is available to her. However, lack of syntactic and discourse management skills leads the student to the simple discourse structuring solution of laying out two independent sentences with repetitive subject referents. There could be

a number of potential ways to improve the structural arrangement and reduce redundancy for (15). A straightforward improvement would be to turn the second sentence into a coordinate clause and omit the anaphoric pronoun *it*. The issue with using a compound sentence would be that it places more or less equal weight of information focus on both the two coordinate clauses and requires them to be closely related to each other. But the semantic relationship between the two sentences in (15) does not seem to be construable as a close one, making the option of compounding the two sentences in (15) somewhat inappropriate. Thus, a more appropriate alternative to edit the discourse syntax would be to use a participle adverbial clause to realise both information integration and syntactic conciseness. Since neither of the two sentences in (15) seems to assume the absolute focus in the local discourse context, it would be fine to turn either of them into the dependent clause and let the other become the matrix clause. Note that a verbless clause instead of the supplementive clause can as well be used to achieve greater economy if the first sentence is to be compressed into the adjunct:

> (15-Restructured) *(Being) a permanent accumulation of one's filtered experience and knowledge, long-term memory holds both our declarative knowledge and procedural knowledge.*

It can be seen that the supplementive clause is useful in integrating into the new construction part of the description and characterisation of a particular entity that only has a marginally attendant relationship with the rest of its predication, as shown in the general discourse semantics of (15). Without unnecessarily breaking the information flow by using two independent sentences, this clause-combining device allows the reader to quickly skim the multiple descriptions of a subject matter under discussion

and to some extent dispense with explicating the latent logical relationship between the adverbial and matrix clause.

The case of (16) differs from (14) and (15) in that the two sentences are not characterisations of the same grammatical subject or textual referent but rather start with two different themes, i.e. *the study of English tense* and *a large number of books and articles*. However, they are similar with respect to the indeterminate between-sentence logical relationship, with the two sentences in (16) jointly contributing to the claim that English tense is a major research topic with prolific publication on it. They do not seem to form a causal relationship but are concomitant representations of each other: English tense being a "main subject" is illustrated and proved by the multitude of works on it. Therefore, the two sentences need to be combined in such a manner as to enhance the coherence of the discourse semantics, as well as syntactic variety. As with the preliminary revising strategy applied to (15), the coherence of (16) can be improved by using clausal coordination of addition (i.e. using coordinator *and*). However, since the additive coordinator *and* in some contexts has the semantic entailment of "and then" or "and so" (Halliday & Matthiessen, 2004, p. 406), its application to (16) may send the message that the relationship between the two clauses is a causal one, which would have violated the student author's original intention for the relationship. Therefore, in order to retain the kind of supplementary contribution of Sentence 2 to the claim made in Sentence 1 and improve its coherence, using a participle adverbial clause would be optimal for the intended purposes, with Sentence 2 rearranged as the supplementive clause to appear after the matrix clause. Here it is the absolute clause rather than the subjectless supplementive clause that should be employed because the subject of the adverbial clause is different from that of the matrix clause (*a large number of books and articles* vs. *the study of English tense*) and

therefore needs to be explicitly presented. To this end, the conjunctive preposition *with* meaning "existence" can be used to introduce the absolute clause, occurring after the matrix clause in sentence final position:

> (16-Restructured) *For linguists and grammarians, the study of English tense has always been one of the main subjects, with a large number of books and articles focussing on this grammatical category in recent years*.

Apart from strengthening the cohesive ties between the two sentences in (16), the revised version also manages to enhance coherence by sending the syntactic signal that the absolute clause introduced by *with* should be holistically interpreted in association with the matrix clause; the ancillary function intended for the second sentence in (16) is now more explicitly exhibited through the elaborative supplementive clause. Note that turning the second sentence into an absolute clause to link with the matrix clause also requires some improvement on lexical cohesion in the revision: to avoid unnecessary repetition of "English tense" which has already occurred in the matrix clause, its hypernym "grammatical category" is used as a cohesive substitute. Thus, it can be seen that enhancing syntactic variety and complexity often necessitates corresponding sophistication at the lexical level.

In summary, the above discourse analysis of student writing extracts has shown that students' underuse of participle adverbial clauses often occurs in relation to certain textual and contextual situations. These include providing historical background, giving multiple characterisations to the same subject entity, and incorporating an attendant clause with a subject other than that of the matrix clause. In addition, underuse is often accompanied by lack of discourse coherence, syntactic monotony, and lexical redundancy resulting from ineffective syntactic arrangement of discourse organisation. The suggested revisions, by restructuring students' own text beyond the sentence level,

have demonstrated how using participle adverbial clauses can improve both syntactic variety and discourse coherence at the same time. Underuse caused by students' underdeveloped awareness of the kind of discourse incoherence resulting from choppy sentences and lexical redundancy as exemplified in (14)-(16) can be greatly remedied by skilful use of participle adverbial clauses. On another note, the enhanced complex sentence in each case has also diversified discourse prosody by duly increasing the number of punctuation units within the sentence and extending the prosodic boundary as demarcated by the full stop. This would allow the reader to focus on and process more interrelated information within a single prosodic break (i.e. each sentence).

6.2.3. Excessive and Aberrant Use of Clause Combination

6.2.3.1. Motivation and Scope

Although corpus-wide overall statistics have shown significant underuse of both finite and participle adverbial clauses in the student corpus, instances of overblown employment of both clause combining devices in student writing have also been observed. That is, the overall picture may be that students produce fragmentary, disjointed sentences in connected discourse that are apparently semantically related and are therefore connectable by clause-combining devices. However, a different scenario permeates the corpus where coherence and understanding are plagued by excessively intertwined use of clause combination. It is not uncommon to find in the student corpus extremely long sentences containing multiple adverbial clauses and other types of dependent structures, sometimes coupled with clause coordination. Such awkward sentences, with complexly entangled hypotaxis and parataxis, would befuddle the reader attempting to locate the focus of the sentence and understand the logical relationships among the many constituents squeezed together. The purpose of this

section is to present and interpret this aspect of clause combination in student writing and draw attention to the negative effects of unduly complex use of clause-combining devices.

The analysis starts with examples of convoluted expression mainly caused by unbridled use of finite clauses and then proceeds to exemplify aberrant employment of participle clauses. But it needs to be noted that many cases of entangled sentences are characterised by excessive use of more than one type of dependent structures, usually with multiple structural types fused in one sentence. Therefore, comments will be made where obstacles to smooth understanding are also attributable to structures other than the two structural types that have been under scrutiny in this chapter.

6.2.3.2. Entangled Finite Dependent Clauses

Entangled use of finite dependent clauses mostly occurs in extremely long sentences, with some instances longer than 100 words. Although involving multiple clauses, these sentences do not usually show much ungrammaticality, with the major problem being fused sentences (or run-on sentences). This reflects Lu's (2010) note that grammaticality does not pose as great a challenge to advanced second language writing as do idiomaticity and pragmaticity. What breaches pragmatic appropriacy in these sentences is a substantial loss of information control and logical clarity among constituent clauses.

From the perspective of information structure, there are just too many turns of information status from topic (or theme) to focus (or rheme) in one sentence, resulting in a multifocal sentence detracting from effective construal. A complex sentence involving convoluted hypotaxis and parataxis along with embedding would be a tough cognitive challenge to the reader's working memory capacity required to process the

interpretation of its meaning. Such a challenge is unlikely to be met since the reader's working memory would not be so powerful as to retain the abundant information and establish logical links between the clauses. That teachers often complain they cannot understand the gist of students' long sentences even after repeated reading is precisely evidence of their working memory being compromised. Confronting a multi-clausal sentence requires readers to constantly revisit the foregoing content as they go through the extremely long sentence.

As previously mentioned, the damage to the interpretation of multi-clausal sentences is not so much derived from sentence length but from the scattering of topic-focus patterns within the sentence. Since each finite clause necessitates a pause, multiple pauses within the sentence would result in the reader's concentration constantly disrupted. The use of adverbial clauses to encode multiple logical relationships in one sentence would further aggravate processing difficulty. The following sentence is taken from the student corpus to illustrate the above characterisation of entangled multi-clausal sentences, with markers of finite adverbial clauses and coordinate clauses highlighted in bold type:

> (17) *Eight interviewees said that* ***although*** *textbooks were often used as a guidebook to direct classroom teaching, materials with close relation with TIE-1 were given excessive attention* ***and*** *as a result, most of the time was used to repeat the "relevant" exercise over and over again,* ***whereas*** *open-ended questions, writing, group discussion, speaking sections were often skipped* ***since*** *these contents were either unexamined or rarely examined in TIE-1.* (Text 02, MDC)

With 67 words in length, this sentence is a complex T-unit containing two *that*-clauses complementing the superordinate predicate verb *said*, both involving the use of finite

adverbial clauses; the second *that*-clause also includes an adverbial clause (the one introduced by *since*) embedded in a higher-level adverbial clause (the one introduced by *whereas*). Thus, this sentence has achieved an enormous grammatical complexity by various measures: length of sentence, length of T-unit, and number of dependent clauses.

However, the problem of the sentence just lies in the fact that it is composed of multiple adverbial clauses: two encoding a concessive or contrastive relation (*although* and *whereas*) and the other encoding a causal one (*since*). Since it can be inferred that including so many adverbial clauses is motivated by the ideational complexity of what the "eight interviewees said", the student's attempt to summarise the eight interviewees' comments in a single effort has contributed to the sentence as it is. Therefore, the reader has to make a couple of pauses to construe each of the logical relationships encoded in the sentence and how they interact with each other to achieve the overall narrative and argumentative purposes intended by the student. But this is no easy task for the reader who may probably have forgotten what has been covered in the first *that*-clause (i.e. the one involving the *although*-clause) by the time she begins to concentrate on the second T-unit (i.e. the one involving the *whereas*-clause). The reader would be forced to go back over the first *that*-clause, several times perhaps, in order to semantically relate the two complement clauses; the semantic relation, as indicated by the use of *and as a result*, should also be a causal one.

The above analysis depicts the communicative ineffectiveness derived from excessive use of finite adverbial clauses encoding defocused semantic relationships within one sentence. The greatest problem with using multiple clause combinations in one sentence is concerned with the extent of cognitive challenge it poses to the reader's working memory in dealing with convoluted logical relationships. However, not just

aberrant multiple combinations of adverbial clauses have been found in the student corpus; there are also instances of other types of hypotactic and paratactic clause combinations which are perceived as deviant. The following two examples are selected which respectively show excessive use of (non-restrictive) relative clauses and postverbal complement clauses within the confine of a sentence, with the relativisers and complementisers highlighted in bold type in the two examples:

> (18) *DMs are used extensively,* ***which*** *do not belong to any of major syntactic categories,* ***which*** *do not indicate objects, events or properties, and* ***whose*** *meanings do not seem to contribute to the conceptual meaning of utterances in* ***which*** *they occur (Wilson, 2000, p.215), once they are moved from the utterance, their semantic meanings can not* [sic] *be changed, but their pragmatic meanings are somewhat different.* (Text 05, MDC)
>
> (19) *However, before deciding* ***that*** *these methods are old-fashioned, counterproductive, and best discarded, Bonnie ought to think about* ***why*** *they have been around so long and* ***what*** *she might learn from them, Bonnie should try to introduce* ***what*** *she considers to be more effective learning strategies, she should also be sensitive to resistance to these strategies, and she should be ready to accept* ***that*** *what works in the West may not be so effective for her Chinese students.* (Text 51, MDC)

These two sentences (65 and 77 words long respectively) also satisfy a range of syntactic complexity standards, including sentence length, T-unit length, and density of dependent clauses. However, they have to be rated as being awkward due to their overuse of constituent clauses. Although it is mainly the use of non-restrictive relative clauses and postverbal complement clauses that contributes to the infelicity, the two examples also involve the use of adverbial and coordinate clauses which make their

internal relations even more entangled. It also needs to be noted that both examples are cases of run-on sentences, which partly explains why the sentences are so extended with convoluted clause combination.

As will be seen, the excessive use of clause combinations attested in (18) and (19) can be argued to have broader linguistic and cross-rhetorical implications. Specifically, the sentence-level overuse of the structure attested in (19) can be related to a corpus-wide overuse. Comparing the frequency of occurrences of non-restrictive relative clauses introduced by *which* between the MDC and PRAC,[38] it has been found that this clause type is significantly overused in the student corpus, as shown in Table 6.3. The analysis included cases of non-restrictive *which*-clauses both with and without a fronted preposition, the former referring to an occurrence such as the following:

> (20) *One final instance may be considered,* ***in which*** *the possible closing launched by the UC question is made explicit in the teacher's launching of the next activity afterward.*

TABLE 6.3

Frequency of Non-Restrictive Which-Clauses in the Two Corpora

	PRAC	MDC	*Log-likelihood*
Basic form (without a fronted preposition)	1,223	2,021	-197.52***
which-clauses following a fronted preposition	130	258	-43.49***
Total	1,353	2,279	-238.21***

Note. Frequency was normalised to instances per million words.

***Difference is significant at the $p < .001$ level.

[38] Compared with other non-restrictive relativisers, *which* is the most frequent non-restrictive relativiser in written genres according to Biber et al. (1999, pp. 610-611), hence only it being focused on in this comparison.

Consistent with the overall overuse of non-restrictive *which*-clauses, student writers use both types of the clause significantly more frequently than expert writers. Although Schleppegrell (1998) found that "all students who used non-restrictive relative clauses had difficulty employing this structure" (p. 196), the present study shows that this clause type is strongly favoured by Chinese students. Since non-restrictive clauses mainly function to add elaborating, descriptive information about the antecedent NP or clause, its prominence in the student corpus may suggest that the student writers are highly keen on using non-restrictive relative clauses to provide additional information and that they perceive its use as a strong feature of academic discourse. This, however, needs further research for better understanding. In addition, the overuse may also point to students' underdeveloped awareness of other structural types with a similar function, including supplementive clauses, appositive NPs, premodification, and the use of separate sentences instead of relative clauses.

Nevertheless, what contributes to the corpus-wide overuse may not properly account for the kind of sentence-level awkwardness seen in (18) and (19). The three consecutive non-restrictive relative clauses in (18) can be interpreted from a cross-rhetorical perspective. In that sentence, three structurally similar constructions are juxtaposed in parallel arrangement: *…which do not*…, *which do not*…, *and whose meanings do not*… By the same token, the three *wh*-word complementisers and the four apparently coordinate clauses starting with the same subject referent (*Bonnie* and *she*) in (19) are arranged in a similar juxtaposing fashion.

Intuitively, parallel juxtaposition does not seem to follow the rhetorical convention of English academic prose, leading one to extrapolate that L1 (i.e. Chinese) rhetorical transfer could be at work giving rise to the attested parallelisms. Since

features of parallelism have also been identified in the use of participle adverbial clauses, explication of how cross-rhetorical influences from Chinese prose traditions may be represented in English academic writing will be provided later in this chapter along with the examination of students' use of participle clauses.

The wider implication brought out from Ex. 19 is indicative of many cases of excessive clause combination, where the sentence has been written in a way characteristic of the many qualities of spoken English syntax. A defining syntactic feature of speech as opposed to formal writing is concerned with the "idea units" in a chunk of discourse (Chafe, 1982, p. 37). As found by Chafe, a typical idea unit in spoken language consisting of a single (finite) clause is about six words in length and the units are either combined through only a small set of conjunctions (notably *and*, *but*, *so*, *because* in order of frequency) or not strung together at all. In addition, spontaneous speech is also characterised by the inclusion of many hypotactically and paratactically related clauses in one sentence,[39] a feature directly related to what Halliday (1987) calls the dynamic style of representing experience.

Therefore, the relatively short clause length and frequent use of basic conjunctions (including zero coordination leading to run-on sentences) to form multi-clausal sentences in (18) and (19) pull their textual features in the direction of spoken registers on the speech-writing continuum. Although previous research has also identified signs of tendency towards an oral style in L2 writing, evidence is culled mostly at the lexical level (e.g. Granger & Rayson, 1998; Paquot, 2010). The findings of this study help to further flesh out the oral tendency by revealing the syntactic similarities between samples of student writing and spontaneous, connected oral production.

[39] Sentence here is considered as having a discourse boundary that can only be prosodically detected.

6.2.3.3. Entangled of Participle Adverbial Clauses

Although corpus-wide data has shown that student writers significantly underuse participle adverbial clauses, especially *-ing* clauses, careful observation of students' use of this clause type has also found sentence-level inflation due to mismanagement. Overuse of participle adverbial clauses and other free adjunct structures (cf. Kortmann, 1991) is exemplified in the following two sentences. The bold typeface indicates the start of all participle adverbial clauses and other adjunct structures whereas the underline highlights the verbs of the matrix clauses:

> (21) *Chapter two reviews the stages of Foreign Language Teaching (FLT),* ***starting*** *from an emphasis on linguistic competence,* ***developing*** *an emphasis on communicative competence, and* ***moving*** *to intercultural communication competence (ICC), then introduces the definitions, components and models of ICC as well as ICC measure instrument: Intercultural Sensitivity Scale (ISS).* (Text 51, MDC)
>
> (22) *This thesis attempts to explore the usage of English posture verb stand,* ***based on*** *the Brown Corpus,* ***contrastive with*** *Chinese cardinal posture verb* li*,* ***based on*** *the CCL Corpus****, revealing*** *patterns of usage with these verbs which are reminiscent of the polysemy and grammaticalization facts,* ***showing*** *functional symptoms typically associated with the posture cohort in terms of frequency, collocation fixedness, tense/aspect-marking, and choice of participants, especially subject.* (Text 55, MDC)

It can be seen that, similar to sentences with excessive use of finite clauses, both (21) and (22) are grammatically very complex in terms of sentence length (49 words and 67 words respectively) and both include a host of finite and non-finite verbs, constituting

all the clauses of the sentence. However, it is precisely the abundance of adjunctive structures that makes the two sentences sound highly awkward. This is because the juxtaposition of so many adjuncts within a single sentence entangles the logical relationships between the adjuncts and the matrix clause as well as between themselves; the reader could be either bewildered or bored by consecutive non-finite and verbless clauses. In terms of sentence-level information structure, it is highly difficult to identify any given-new distribution of idea flow, as if all the non-finite and verbless structures after the matrix clause contribute a different portion of new information to the sentence. The fragmentary adjunctive structures chained one after another within the same sentence would disorient the reader as to what to focus on and expect from the stretched rhematic layer.

In interpreting the overflow of participle adverbial clauses and other adjunct structures within a single sentence, considerations about both discourse management and cross-rhetorical transfer need to be entertained. From a discourse management perspective, it is apparent that the students have failed to rein in control over the huge amount of information and ideas that they see as relevant to the matrix clause and to one another, only finishing off when they feel everything has been covered. They have also failed to adequately appreciate the textual functions of adjunct structures to supplement, explain, comment on, or otherwise qualify the matrix clause. These supplementation, explanation, and comments play a fundamentally subordinate role in relation to the matrix clause and so only one such structure accompanying the matrix clause to appear in a sentence would be optimal.

The slew of adjunct structures used in a chain could also be a consequence of students' conscious attempts to avoid using too many finite structures and thus to achieve an integrated style: if the adjunct structures are to be restored into finite clauses,

they either combine through clausal coordination or appear as independent sentences with a repeated grammatical subject or topical theme, neither of which achieve the kind of stylistic economy of adjunct structures. Therefore, the strategy can be seen as a negotiated product of making stylistic choices in general and thematic choices in particular.

As with the overuse of finite dependent clauses seen in (18) and (19), the reason why students also use an undue number of participle clauses in a sentence can be also approached from the perspective of cross-rhetorical transfer. The heavy parallelism so far attested may have been derived from the rhetorical pattern of traditional Chinese writing, which is strongly rooted in the rules laid down for composing the "eight-legged essay". This examination-based essay form requires the use of parallel sentence structures as an important syntactic device for writing up the vast majority of the eight legs, or sections, of the essay. Influences of this grand tradition of using parallel structures at the phrasal, clausal, and discoursal levels in the Chinese rhetorical tradition are still widely traceable in modern Chinese rhetoric, especially in institutional and bureaucratic texts. Consider the following excerpt from the 2015 "Government Work Report" delivered by Chinese Premier Li Keqiang (The State Council, 2015):

> (23) ***We worked to strengthen*** *employment and social security.* ***We improved*** *the policies to stimulate employment and initiated the scheme to help college students and graduates to start businesses, ensuring a steady increase in the employment of college graduates.* ***We unified*** *the basic pension systems for rural residents and non-working urban residents and increased basic pension benefits for enterprise retirees by a further 10%.*

As indicated by the bold highlights, the recurrent use of *we* plus a past tense verb across the short excerpt, which is actually characteristic of the rhetorical strategy of the whole

report, embodies one way in which parallelism contributes to textual development of the report. The highly valued text-generating formulae of ritualised forms of Chinese discourse may lend itself to explaining the intensive paralleling of both finite dependent clauses and non-finite adjunctive structures attested in Chinese student academic writing. Just as the lexicogrammatical features of L2 learners' L1 may be negatively or positively transferred to the language being learnt, so is it true for features of higher-level discourse grammar and style. Furthermore, students' keen experimentation with parallelism does not conform to the stylistic requirement of parallel arrangement of English writing in which all the constituents in the series are usually of the same structural type. The parallel structures in (21) and (22), by contrast, are a mix of different types of adjunctive structures: both *-ing clauses* and finite clauses in (21) and an even broader mix of clause types in (22).

From another perspective, the chaining of participle adverbial clauses and adjunctive structures within a sentence may be seen in the same light that explains the convoluted combination of multiple finite clauses, both subordinate and coordinate. That is, it may again derive from implications of Halliday's (1987) dynamic style.

To sum up, it has been shown that although Chinese student writers underuse participle adverbial clauses in overall quantitative terms, instances have been found where the said structures show stylistically disturbing overuse among local discourse chunks, i.e. within aberrantly long sentence. Factors concerning both discourse-syntax management and cross-rhetorical influences have been taken up for explaining the overuse. Although the overall underuse of participle clauses is a general reflection of students' underdeveloped awareness of their functions to integrate contextually relevant information and enrich structural variety, the attested excessive use indicates

students' lack of discourse-syntactic resources to make better thematic choices and avoid awkward parallelism.

6.2.4. Summary of Overuse of Clause Combination

The sporadic, here-and-there overuse of both finite and participle adverbial clauses at the sentence level observed in the student corpus helps illuminate issues regarding students' writing processes and the selection of measures for validly representing the syntactic complexity of L2 academic writing. The inclusion of multiple hypotactically and paratactically related clauses within one sentence indicates the difficulty with which student writers try to "juggle" multiple ideas, to use the term borrowed from Barton, Halter, McGee, and McNeilley (1998). In other words, it is the semantic complexity perceived by student writers that contributes to the surface-level structural complications. Sometimes it may be the case that students perceive the message encoded in the matrix clause to be inadequate, something that could have prompted them to add many details to supplement the basic superordinate idea; this is reflected in the use of multiple parallel clauses following the relatively short superordinate clause in (18), (21), and (22). In the case of (17), the student could not help explicating in one single sentence all the contrastive and causal relationships concerned with the complicated tension in teaching materials application.

The above observations seem to indicate that convoluted clause combination is driven by students' keen but misbegotten belief that this strategy contributes to grammatical complexity and hence writing quality. However, they may not have been made aware of the amount of clause combination and features of grammatical complexity at the sentence level in professional academic writing. Rather than

containing multiple clauses, the complexity of the typical sentence in expert writing rests upon the extent to which structural integration (Chafe, 1982, p. 37) is realised.

Frequency effects (Ellis, 2002) in relation to the normal density of clauses within a sentence (or a T-unit), as would be expected from enormous exposure to published academic texts, do not appear to be effectively at work in sensitising students to appropriate clausal elaboration. On the contrary, it is the kind of mechanisms governing online oral production, i.e. spontaneous emergence of new clauses conditioned by the corresponding idea flow, which is held to account for the somewhat awkward multi-clausal structure. This may lend some support to understanding students' reluctance to wind up their sentence when having to express semantically complex relationships. Since sentences as exemplified above are either awkward or pose obstacles to smooth and clear logical presentation, they do not easily provide much room for revision without major revamping (Barton et al., 1998; Gosden, 1992). Revision usually requires the multiple clauses constituting a long sentence to be rewritten as separate sentences or using more integrated structural types, such as nominalised verb phrases and adjunctive structures.

With respect to the selection of effective syntactic complexity measures, it can be seen that just counting the number of clauses used in a production unit—a sentence in particular—to gauge grammatical complexity may sometimes ignore instances of students' texts where overuse of clauses blocks smooth comprehension. Although students have been found to underuse both finite and non-finite adverbial clauses overall, their writing often reflects attributes of oral production, e.g. the use of multiple adverbial clauses to weave together experiential representations. Therefore, it can be said that novice student writers use adverbial clauses somewhat differently from expert writers, in that they are used not so much to encode semantic relationships between

ideationally complex statements (mostly achieved through the use of complex noun phrases) as to "indicate the knowledge base for their assertions" (Schleppegrell, 1996, p. 271) and to mediate an extended succession of information flow. Instances of multi-clausal sentences with entangled between-clause relationships found in the student corpus to some extent justify recent criticisms directed at using measures centred round clausal density to gauge grammatical complexity (e.g. Biber et al., 2011).

However, it needs to be clarified that speech-like clause combination does not in fact constitute a major tendency in student writing, only occurring at irregular intervals and easily overlooked by mean score comparisons. Therefore, there seems to be a tension in recognising the effectiveness of measures of clausal density: its ability to point to important underuse of certain types of circumstantial adverbial clauses as a discourse feature of student writing versus its easily misinterpreted emphasis on clause combination which, when mismanaged and overused at the sentence level, can result in entangled meaning relationships and an oral tone of writing. Students who write in deliberate response to the clausal density measures would probably be prone to overuse clause combination at the expense of meaning relation clarity and structural variety. Therefore, it can be argued that although clausal density measures should still deserve a prominent position in gauging the grammatical complexity of clause combination, its interpretation without considering the semantically driven effectiveness of how dependent clauses are used in specific discourse contexts would be detrimental to the measures' potentially positive washback effect on students' writing development. Investigations into the quality, not just the quantity, of students' use of dependent clauses in relation to semantically-benchmarked evaluations of their effectiveness need to be incorporated in any measurement framework for gauging clausal complexity.

6.3. Summary of Chapter

In response to the findings that student writers use significantly fewer finite dependent clauses and verb phrases than expert writers based on the results of comparing syntactic complexity measures, this chapter has examined the quantity and quality of specific types of both finite and participle adverbial clauses. Finite circumstantial adverbial clauses of three semantic categories (i.e. reason, concession, and condition) and participle adverbial clauses are included for frequency comparison. It has been found that student writers significantly underuse both finite adverbial clauses and participle adverbial clauses overall in comparison with expert writers, with different categories of the two types of adverbial clauses making varying contributions to global differences.

Specifically, student writers use significantly fewer causal and concessive clauses (marked by the canonical conjunctions *because* and *although* respectively) but significantly more conditional clauses (marked by *if*), indicating that student academic writing as a developing genre is still weak in adequately providing explicit account, justification, rebuttal, and countered-rebuttal to support or problematise claims and assertions made in the ongoing discourse. Thus, a complex argument web characteristic of high-quality academic writing fails to be weaved in student writing. Textual evidence has been identified to illustrate discourse chunks where between-sentence causal and concessive relationships are left latent and can be made explicit through clause-combining devices. On the other hand, students also significantly underuse participle adverbial clauses overall than expert writers but the gap is mainly attributed to inadequate use of *-ing* clauses, in all positions of the sentence. Textual analysis has also revealed discourse circumstances where it would have been preferable to employ participle adverbial clauses for better coherence and information flow. The

underrepresentation of this clause type, which is unique to the written registers, points to students' underdeveloped awareness of the prominence of particular structural types in academic discourse.

However, against the panoramic landscape of underuse of both finite and participle adverbial clauses, overuse of clause combination at the sentence level has been found in the student corpus, leading to entangled between-clause meaning relationships and awkward structural arrangement. The erratically extended chaining of dependent clauses has been seen as related to situations where students need to juggle multiple complex ideas which they perceive to be logically connected as a whole. Explanations from a cross-rhetorical perspective have also been attempted at, attributing instances of strong parallelism to rhetorical traditions valued in both ancient and modern Chinese prose. In addition, considering the varied linguistic features on the speech-writing continuum, the process in producing the multi-clausal sentences turns out to resemble that for speech, indicating that student writers tend to be unconsciously conditioned to a process leading to a style featuring spontaneous uttering of narrative-like, dynamically structured ideas.

The findings also have implications for updating theoretical and empirical understandings of subordination measures in gauging the syntactic complexity of L2 academic writing. With respect to the developmental prediction that L2 writers tend to produce fewer dependent clauses along with age or proficiency level (Ortega, 2003), findings from this study have shown that the use of adverbial clauses—both finite and participial—is actually an area where student writers lag significantly behind expert writers, at least in terms of adverbial clauses of certain semantic and structural types. This suggests that although advanced student writers may have broken away from a highly oral style associated with frequent use of conjunctions, they face a different kind

of challenge from the cognitively laborious job of encoding complex causal, contrastive, concessive relationships for which circumstantial adverbial clauses serve as a major means. In other words, when it comes to using adverbial clauses in conformity to the discourse functions and level of semantic force required of academic writing (Schleppegrell, 1996), students tend to take an avoidance strategy. In this sense, the lack of syntactic complexity, in terms of combining clauses to encode meaning relationships at least, is directly linked with students' underdeveloped domain-specific disciplinary and epistemological maturity and sophistication.

Results of clause density measures indicating students' underuse of certain types of clauses in this study should only be taken as reflecting a general generic tendency and careful observation is needed to better understand how this global tendency can or cannot fit into the local evidence of clause linking. Finally, participle adverbial clauses which are traditionally incorporated in measures of verb phrase density should be necessarily used for clause measures.

CHAPTER SEVEN

CONCLUSION

This chapter summarises the major findings of the study, discusses their theoretical contributions and pedagogical implications, reflects on some of the principal limitations of the study, and proposes some recommendations for future research.

7.1. Summary of Major Findings

This study has examined the syntactic complexity of EFL academic writing at the postgraduate MA level, providing answers to three overarching research questions raised in Chapter 1:

Research Question 1: Which syntactic complexity measures show significant differences between EFL academic writing and expert academic writing?

Research Question 2: What are the grammatical features of noun phrase complexity where student writers lag behind expert writers?

Research Question 3: What are the grammatical features of clause combination where student writers lag behind expert writers?

With a contrastive corpus-based approach, these three questions have been investigated in three corresponding chapters. For this purpose, two corpora were built for the study, one of EFL student academic writing and the other of expert academic writing. The student corpus is composed of 70 MA dissertations written by Chinese-speaking students studying at universities on the Chinese mainland while the contrastive expert corpus comprises 129 research articles published in high-impact journals. Both corpora are in the broad academic discipline of applied linguistics. The rationale behind comparing student writing with professional expert writing was to identify the areas of

syntactic complexity where student writers lag behind expert writers, which can be effectively taken up for focused EAP instruction. The major findings of each chapter are summarised below.

The study first employed the L2 Syntactic Complexity Analyzer to automatically compute its 14 syntactic complexity measures for the two corpora. It has been found that syntactic complexity in the student corpus is significantly weaker than in the published corpus on all but two measures: T-units per sentence (T/S) and clauses per sentence (C/S). In terms of the three major dimensions of syntactic complexity proposed in L2 writing research, both subordination-based complexity measures and measures involving noun phrase complexity have shown significant differences. However, the two corpora do not show any significant differences on measures of clausal coordination. In other words, student writers need improvement in constructing both types of syntactic complexity, except for the use of coordinate clauses, which is itself not a prominent feature of academic discourse.

One aspect of the results that is somewhat surprising about the results is that both the phrasal and clausal dimensions of syntactic complexity of student writing are weaker compared with expert writing. This needs to be related to the developmental prediction (Ortega, 2003) that student writing would show weaker noun phrase complexity but use more dependent clauses which are said to characterise informal speech; however, the present study does not seem to support this prediction. A possible explanation for this could be that the student writers involved in this study have largely moved away from features of orality associated with the use of consecutive dependent clauses (e.g. with such subordinators as *because*, *if*, *when*, *what*, or *how*) in a long segment of utterance. Instead, they may have gone to another kind of extreme in combining clauses, i.e. using short clauses that are not adequately combined. Thus,

student writers may have been challenged both in packaging factually-based information into layered noun phrases and in constructing effective logical relations among clauses.

The study went on to examine various structural and usage-based aspects of noun phrase complexity in student academic writing, drawing on theoretical and empirical insights derived from functional and corpus linguistics. The analyses were based on samples of complex noun phrases expanded from a selected set of head nouns. Among the findings, the following general patterns merit summarisation.

Regarding the variety of postmodifying structural types, the following findings stand out. It has been found that student writing shows a slightly higher proportion of clausal postmodifiers versus postmodification by preposition phrases. Among prepositional postmodifiers, student writers use more preposition phrases headed by *of* than by other prepositions. Student writing also shows a greater use of postmodification by finite clauses than non-finite clauses. These findings suggest that student writers appear to prefer canonical postmodifying structures which are hypothesised to develop earlier in the second language and are more characteristic of informal speech and academic writing (Biber et al., 2011). With a strong reliance on these early acquired postmodifying devices, student writers have not made adequate attempts at other types of postmodifiers that are more efficient and flexible in compressing information.

Another explanation for students' less frequent employment of other prepositions and non-finite relative clauses in postmodification could be related to the relatively low frequency of prepositions other than *of* and the lack of distinct relativisers (e.g. *that* and *which*) in non-finite relative clauses. Explicit instruction is needed to direct students' attention to specific noun-preposition combinations other than those involving *of* and the wide spectrum of meanings carried by these combinations.

Similarly, it would be rewarding to guide students to take heed of expert writers' choice between finite and non-finite relative clauses and their semantic and functional correlates in discourse. However, these generalised results should take into account the semantics of specific head nouns in the selection of postmodifier types, with the postmodifying pattern for some nouns largely deviating from overall distribution.

The examination of the complexity of multiple postmodification has found that student writers produce shorter and fewer levels of consecutive postmodification in constructing complex noun phrases than expert writers. These results help to explain students' significantly shorter mean clause length and significantly fewer complex nominals. Believed by some to be the latest developed and most distinct postmodifying device in academic writing (e.g. Biber et al., 2011), postmodification by consecutive preposition phrases realises the most extreme form of information condensation. Turning multiple verb-framed clauses into a linear string of phrases allows the writer to make descriptions, delimitations, and definitions as complete and detailed as possible; economy of expression is among the strongest drivers. Thus, the findings suggest that students either are not able to provide detailed information due to their lack of disciplinary knowledge and engagement, or they have not been made aware of the prominence of this discourse feature and relied on clauses for detailed explication. However, multiple postmodification in expert academic writing need not be exaggerated and is found to centre around two consecutive postmodifiers, only seeing a small proportion of exceedingly complex instances. That being said, what needs to be emphasised is rich information packing within each individual level of postmodifier through the use of one or more premodifiers.

Putting the use of CNPs into meaningful context, the study further examined the use of complex noun phrases in clause subject position and the scope of meanings

assigned to them via the choice of predicate vocabulary. The results show that student writers use fewer CNP subjects than expert writers in general, with CNPs headed by *analysis* showing the greatest gap between corpora; and the two corpora also exhibit a significant difference in the length of subject CNPs headed by all three head nouns. Lastly, results of the predicative vocabulary for CNP subjects show that transitivity-wise student writers use a smaller variety of transitive verbs than expert writers, as well as fewer instances of the copula + adjective + (optional) complement construction. In terms of lexical richness of the predicative vocabulary, student writers use a smaller range of vocabulary to predicate all three sets of CNP subjects and the lexical items they use show a weaker degree of lexical sophistication. Students' insufficient and less elaborated use of CNP subjects can be illustrated in terms of Vande Kopple's (1994) characterisations of the discourse functions of long subject NPs and three sources of pressures facing science writers to produce long subjects.

With the most prominent function of long subjects being to integrate both given and new information together with their evaluation, student writers do not seem to have developed a keen awareness of organising complex information in the CNP subject. They have neither been pushed to produce long subjects by the three exigent pressures of precision, economy, and efficiency and progressiveness in constructing claims that inevitably haunt expert writers. The findings about the lexical choices made to predicate CNP subjects indicate that using highly abstract subjects poses great difficulty for students to assign clear, and usually metaphorical meanings to the subjects. Given that being able to skilfully construe and gloss abstract subjects represents the most advanced form of grammatical metaphoricity, this ability calls for students' linguistic growth well beyond just the lexis.

The last major part of the study examined the complexity of clause combination in student writing, focussing on the use of finite adverbial clauses of three semantic categories (i.e. reason, condition, and concession) and participle adverbial/adjunctive clauses in different syntactic positions. Results show that student writers significantly underuse both finite adverbial clauses and participle adverbial clauses overall in comparison with expert writers, with the differences varying across semantic and structural categories. Specifically, student writers use significantly fewer causal and concessive clauses but significantly more conditional clauses, indicating inadequacy in providing explicit accounts, justifications, rebuttals, and counter-rebuttals to support or problematise claims and assertions made in the ongoing discourse. This interpretation has also been substantiated by textual analysis, which has found instances where claims need to be enhanced by using adverbial clause combination and where this enhancement is hidden among independent clauses. On the other hand, students also use significantly fewer participle adverbial clauses overall than expert writers but the gap is mainly attributed to inadequate use of *ing*-clauses, in all positions of the sentence. Textual analysis has revealed discourse circumstances where it would have been preferable to employ participle clauses for better cohesion and coherence. Observation has also been made on instances of excessive use of both finite and participle clause combination in student writing, which results in entangled between-clause meaning relationships and awkward structural arrangement. Explanations from the perspectives of cross-rhetorical analyses and awareness of speech-writing distinction were invoked for the erratically combined clauses.

7.2. Theoretical Contributions and Pedagogical Implications

The findings of this study have both theoretical contributions and pedagogical implications for the study and teaching of EFL academic writing. Theoretically, the study contributes to the call for syntactic complexity research to be conducted in association with its discourse-functional and usage-based underpinnings; at the same time, the relative importance of noun phrase complexity and the complexity of clause combination has also gained updated insights from the findings. Pedagogically, the study has revealed areas of syntactic complexity where student writers lag behind expert writers, which can be taken up for explicit language-focused instruction by EAP practitioners. The specific theoretical contributions and pedagogical implications are presented below.

7.2.1. Theoretical Contributions

Findings from the study have three important theoretical implications for L2 syntactic complexity research.

First, this study pushes the boundary of L2 syntactic complexity research to be contextualised in its discourse-functional and usage-based underpinnings. Syntactic complexity measures—mainly used for the sake of tracking L2 proficiency development—have been largely based on crudely defined grammatical categories and motivated on a formal, structural basis (Ortega, 2015). While this study continues this structural line by employing syntactic complexity measures derived from it, substantial progress has been made in seeking functionally and discourse-semantically driven selection of complexity indexes and interpretations. This is particularly needed at the most advanced levels of instructed L2 development, e.g. at the level of postgraduate research writing (Ryshina-Pankova, 2015). This motivation has given rise to some

methodological innovations as have been seen in this study. Specifically, by establishing some of the most distinctive grammatical features of academic discourse as informed by contributions from register variation studies in functional linguistics and corpus linguistics (e.g. Biber et al., 1999, 2011; Halliday & Martin, 1993), the study has focused on various aspects of noun phrase complexity and clause combination as two important dimensions of the syntactic complexity of academic writing. Thus, this study represents a major step forward for L2 syntactic complexity research by attempting to break down complexity measures according to functional and usage-based entailments.

Second, this study focuses on detailed examination of noun phrase complexity in its own right as an important component of multi-dimensional syntactic complexity. Although characterised as the most distinctive grammatical feature of academic discourse, previous research on L2 syntactic complexity has not given it adequate attention it deserves. Commonly used measures such as complex nominals per clause and complex nominals per T-unit tap into distinct nominal structures of which the use of complex noun phrases is a part. Therefore, this study has teased out complex noun phrases from such measures and carefully examined its structurally and functionally significant component of postmodification. Such a treatment makes it possible to scrutinise how the complexity in NP postmodification contributes to the complexification of clause elements (e.g. subject, object, and adverbial) and what prevents students' effective use of postmodification.

Moreover, the generically significant feature of complex noun phrases used in subject position has been investigated, particularly in relation to their meaning-generating process instigated by lexical choices in the predicate. Although previous studies have emphasised using inanimate, abstract subjects in academic writing and the

use of nominalisation as a central feature of grammatical metaphor, they have either only focused on short NP subjects together with their predicate vocabulary (Callies, 2013) or complex nominalisation-induced constructions without their meaning-making roles in the clause (Byrnes, 2009). Such being the case, this study has contributed to addressing this gap by examining grammatical metaphor in student writing from the perspective of how it creates difficulty for student writers and how grammatical metaphoricity can be better realised in the semantic links among clause elements.

Third, the study has shown that the view that clause combination is a prominent feature of spontaneous speech and should therefore be played down in writing instruction and measurement (e.g. Biber et al., 2011) needs to be reassessed. The developmental hypothesis (Ortega, 2003, p. 514) predicts that advanced proficiency groups would draw on complexification at the phrasal, rather than the clausal level, with the latter characterised as a feature of immature writing. However, this study does not seem to support this prediction in that expert writers—who should be necessarily conceptualised as a high proficiency group—have shown to use a significantly greater amount of subordination than student writers. Both the ratio-based syntactic complexity measures of overall subordination such as clauses per T-unit (see Table 4.3) and the frequency-based finite and non-finite adverbial clause indexes have shown this pattern (see Table 6.1 and 6.2). At the same time, expert writers also capitalise significantly more on noun phrase complexity than students. Therefore, the concurrent rather than curvilinear relationship between these two theoretically contradictory aspects of syntactic complexity calls into question the capacity of the developmental prediction to explain syntactic complexity development for highly advanced L2 writers.

It seems that students' performances on creating noun phrase complexity and clause combination as shown in this study take place on two different meaning-making

levels, not just a simple either-or issue. The theoretical insights offered by this study may contend that after a probable increase in noun phrase complexity and drop in subordinate structures through early advanced stages of literacy development (say, undergraduate years), L2 writers need continue to seek further development in these two modes of meaning-making resources. Although students may have refrained from a speech-like pattern in clause chaining and learnt to pack information as phrases, they may have overlooked the necessity of establishing logical links between densely structured statements. Therefore, clause combination for the purpose of academic writing should be conceived of as conditioned by a different set of cognitive processes and semantic forces from that for conversational purposes. This makes legitimate the inclusion of subordination-based complexity for measuring and teaching syntactic properties in the EAP profession. Future research should direct more attention on how the interaction between clause combination (inter-clausal) and noun phrase complexity (clause-internal) impacts (in)effective academic writing.

However, clause combination in student writing has shown a more complicated picture than mere overall underuse, i.e. it is also characterised by aberrantly excessive overuse of an assortment of clause types, sometimes in paragraph-length sentences (see examples in Section 6.3.3.2). Theoretically, this can be either ascribed to underdeveloped awareness of the differences of clause combination between conversational and academic registers on the one hand and cross-rhetorical influence from features of conventionalised clause arrangement for text development tradition in students' L1 on the other. In this study, the identified features of clause combination in EFL writing needs to take into account the literacy tradition of Chinese. Since there is little rule constraining clause arrangement in written Chinese texts of both ancient and modern times—rather, clauses are largely connected on semantic grounds—

students may have unconsciously translated this potential of using multiple clauses in their English writing. This explanation supports the cross-rhetorical transfer prediction proposed by Ortega (2003) to join force with syntactic complexity research to better account for the role of L1 rhetorical features in the syntactic features of L2 writing.

7.2.2. Pedagogical Implications

With the ultimate purpose of this study being able to provide pedagogical implications for advanced EFL academic writing instruction, a number of specific implications can be drawn from the findings of the study, in relation to constructing and unpacking noun phrase complexity and clause combination, as well as possibilities for data-driven learning.

7.2.2.1. Pedagogical Implications for Noun Phrase Complexity

First, in terms of deploying grammatical resources for postmodification, it would be necessary for students to be aware of the possibility of using structural types other than the canonical lexicogrammatical devices. In particular, the findings show that among relative clauses students should make more attempts at non-finite relative clauses as an alternative to their finite counterpart. Students need to be shown that non-finite relative clauses can be especially useful when it is not necessary for the writer to explicate the tense and modality of the predicative process, for example:

> (1) *Lesson 1 had learners formulate* ***a rule relating*** **se** ***to impersonal expressions like (1b)*** *based on a reviewer s description of two Costa Rican restaurants…*

Using participle relative clauses helps to create communicative economy and focuses the reader's attention on the important message only. While the decision between

participle relative clauses and their finite counterparts in specific contexts is a subtle one to make, perhaps requiring native-like intuition (cf. Granger, 1997), student writers can at least be encouraged to experiment with using these devices to achieve syntactic brevity and conciseness. This can be implemented by asking students to evaluate their own finite relative clauses and find opportunities where they can be abbreviated as non-finite structures. Regarding using prepositional postmodification, preposition phrases headed by prepositions other than *of* can serve as useful complementarities as postmodifiers. Apart from using canonical prepositions for certain de-adjectival or de-verbal nouns (e.g. the use of *on* with *dependence* or *between* with *relation*), other prepositions can also be flexibly selected based on specific context-dependent spatial or temporal relations (both concrete and abstract). Take for example prepositional postmodification for *development* as head noun. In the academic sub-corpus of COCA, besides the most common preposition *of* for the head noun, preposition phrases headed by *in* and *for* also occur with considerable frequencies, such as

(2) *literacy development **in** young children*

(3) *curriculum development **for** music education*.

In fact, these two prepositions are noted by Biber and Gray (2016, p. 117) as particularly common in addition to *of* in academic discourse. Moreover, there is also a range of grammaticalised complex prepositions—e.g. *in terms of* in the following example—that are commonly used for postmodification purposes, which should also be an important focus of pedagogy:

(4) *HB's development in terms of reading skills*.

Given the highly abstract semantic relations between nouns and preposition phrases, it would be helpful for teachers to help students unpack these relations by restoring them to their prototypical clausal forms. Students can be shown the range of meanings

expressed by prepositions other than *of* in expert academic writing, which have gained enormous semantic variety and flexibility over the past few centuries (Biber & Gray, 2011, p. 241).

Second, pedagogical efforts need to be made with respect to teaching about the complexity of multiple postmodification in the complex noun phrase. Both EAP and domain-specific teachers can prompt students to explicate abstract processes, entities, and attributes as explicitly as possible via postmodification, which usually needs multiple levels of postmodifiers. However, it is necessary for teachers to point out that multiple postmodification with four or more consecutive modifiers is uncommon and too much of this may undermine effective comprehension. Students can be taught to draw on contextual cues to enhance noun phrase complexity. Specifically, wherever students' use of noun phrases lacks adequate clarity, teachers can ask students to clarify the temporal, spatial, relational, and logical details that can be packed into complex noun phrases. This of course requires both teachers and students to closely engage with the research domain or subject matter being written about, with the necessity to have much recourse to intertextual sources through enhanced literature review.

Third, the discourse features of using complex noun phrases in subject position merit instructional efforts for students to better understand how the complexity of noun phrases is driven by discourse needs. This can be taught in conjunction with the information structure of academic discourse, in which contextual information is often embedded in the subject position. A further contextual approach to enhancing students' use of complex subjects can be based on prompting them to empathise with the kinds of pressure often accompanying expert writers (see 2.4.2.3). Moreover, although a challenging task, students also need learn to broaden the semantic scope of complex subjects to be realised by a greater variety of predicate vocabulary, particularly

transitive verbs of greater lexical richness as well as prepositions and adjectives marking temporal and spatial relations. This process is crucial for gradually orienting students to more synoptic, metaphorical, and paradigmatic ways of construing intellectual experience.

7.2.2.2. Implications for Clause Combination

This study sheds crucial light on areas of student writing where clause combination—finite and participle adverbial clauses in particular—can be improved to create better textual cohesion and coherence and to enrich and diversify syntactic style. Now that it has been found that student writers use significantly less clause combination than expert writers overall, EAP instructors should reconsider its importance in academic writing and focus on teaching features of clause combination that are unique to this register.

This study raises the need to teach the contextual and textual significance of using adverbial clauses.

Since the production of finite adverbial clauses in and of themselves is minimally syntactically challenging—just using a subordinator will do most of the trick—the real difficulty for student writers would be at the level of discourse semantics. That is to say, teachers need to raise students' awareness of the necessity of using circumstantial adverbial clauses to explicate such argument-sensitive elements as reason, justification, concession, and rebuttals, given that *because*- and *although*-clauses are significantly underrepresented in student writing. However, this would not come about easily as it may seem, since being able to do this hinges upon the writer's level of expertise in the research area being written about and strong critical thinking capacities. Therefore, teaching this type of clause combination needs to work in tandem

with the pedagogy of argumentative structures and critical thinking skills. Teaching in this respect can be based models of argument structure (e.g. Crammond, 1998) which emphasises the presence of data, claim, warrant, rebuttal, and counter-rebuttal as building blocks of good argumentation; among other argument building devices (e.g. disjuncts), these elements rely on using circumstantial adverbial clauses to a great degree. But it would be mandatory for teachers to show students the prominent difference in semantic force of inter-clausal logical relations between academic writing and speech, which is mainly achieved through the complexification of the noun phrase in academic writing.

As for participle adverbial clauses on the other hand, the data shows that students' problem mainly lies in a lack of explicit marking of thematically-related statements that could have been linked up through *-ing* or *-ed* clauses. Therefore, instructional efforts in this regard should be focused on raising students' awareness of using these clause types for encoding statements that serve to provide a background for the main claim or to present supplementive and attendant characterisations for the same grammatical subject. The former aim is typically achieved with a sentence-initial or medial participle clause (i.e. before the main clause verb) while the latter with final-position participle clauses (i.e. after the main clause). However, despite students' overall underuse of adverbial clauses, the data has also shown cases of intermittent overuse of clause combination within a sentence, to the extent that clauses get entangled together and their meaning relations are almost lost. These aberrantly combined multiple clauses in student writing severely undermines effective comprehension and teachers should help students avoid this kind of clause combination characteristic of long stretches of spontaneous speech.

7.2.2.3. Implications for Data-drive Language Learning

Finally, corpus-based data-driven learning and discovery learning principles (Boulton, 2010; Kettemann & Marko, 2002; Warren, 2016) can be employed to facilitate the instruction of both noun phrase complexity and clause combination. This can be carried out either through corpus-based presentation and exemplification of these features in expert writing by searching for key discourse-syntactic markers or through discourse analysis of students' written product as shown in this study. To do the former, teachers can use both publicly available corpora of published academic writing in general (such as COCA-academic) and self-compiled corpora of academic writing in a specific discipline (such as the expert corpus used in this study). Students can be introduced to techniques for data-driven learning whereby they can learn to observe the syntactic and discourse features of certain types of clause combination used by expert writers by accessing concordance lines containing them. For example, features of noun phrase complexity can be observed from concordance lines extracted by searching for colligations of specific head nouns co-occurring with prepositions (e.g. *analysis of*) or collocational frameworks associated with a wider scope of prepositional postmodification (e.g. *of the*). Similarly, students can search for subordinate conjunctions typical of different semantic categories (e.g. reason, concession, and contrast) and observe the complexity of logical relations encoded among clauses and how they help build claims and arguments. The same sets of corpus searches can also be performed with corpora of students' own writing to allow them to identify syntactic weaknesses common to EFL academic writing in general and make comparisons with expert writing. In the process of data-driven learning, it would be necessary for teachers to explicate the functional and usage-based significance of lexicogrammatical decision making and how the failure to achieve this significance has led to ineffective writing.

7.3. Limitations of the Study

Despite the study's pedagogical and theoretical implications as presented above, a number of limitations need to be acknowledged.

First, although claiming to be an investigation of EFL academic writing, this study only takes as data academic texts written by Chinese mainland students, leaving out EFL writing by students of other cultural and linguistic backgrounds. Therefore, this limited data selection raises the problem of generalisability of the findings of the study. Given a strong association between syntactic complexity and L1 found in previous research (Lu & Ai, 2015) and informed by the cross-rhetorical transfer hypothesis (Ortega, 2003), the findings of this study may not be generalisable to syntactic features of texts produced by students of other L1 backgrounds. In addition, it is also not known if the findings can apply to the writing of Chinese-speaking students studying in an EMI educational context (e.g. Hong Kong), as syntactic complexity may be greatly enhanced by the language of instruction. Another data-related constraint has to do with the issue of disciplinarity, i.e. this study limits itself to student writing in applied linguistics which is somewhere in the humanities and social sciences, leaving other disciplines, especially those in the natural and engineering sciences unaccounted for.

Second, although the notion of genre has been emphasised as a crucial factor affecting the lexicogrammatical landscape of overall discourse, this study has made no distinctions among the various sub-genres, or rhetorical sections of either the dissertation or research article as providing differential motivations for linguistic choices. Since different rhetorical sections vary greatly in communicative purpose, the lexicogrammar used to realise each section may necessarily be different. For example, there could be more passivisation in methodological descriptions than in introductory

sections, which can give rise to increased use of complex noun phrase subjects. Similarly, the literature review segment may see many complement clauses used after a reporting verb, possibly resulting in an upturn of subordination measures. Therefore, future research should examine how well students' use of syntactic features fits in with the rhetorical needs of each sub-genre.

Third, the data sampled for the detailed study of noun phrase complexity is only limited to noun phrases expanded from a few head nouns. Although allowing for economy of analysis and observation of head-specific variation, the small set of noun phrases may not adequately represent noun phrase complexity in its entirety. A comprehensive profiling of noun phrase complexity is needed by coding all instances of noun phrases from a much smaller dataset. By the same token, the study of clause combination is only based on adverbial clauses connected by a limited number of subordinators: only *because*, *if*, and *although* for finite adverbial clauses, and only supplementive and *with*-headed clauses for participle adverbial clauses (i.e. excluding participle clauses headed by conjunctions). The inclusion of all the other types of adverbial clauses would have made the description more comprehensive and complete. Another weakness of this study is its lack of ethnographical insights either from students on their expectations and challenges concerning syntactic complexity or from EAP writing teachers on how syntactic complexity affects their evaluation of meaning-making appropriacy.

Lastly, for all the emphasis on creating syntactic complexity, the study lacks a critical dimension in terms of how such complex phenomena as heavy noun modification and clause combination may undermine clarity and effectiveness of communication. As much has been lost in compressing clauses into phrases, including tense/aspect, mood, agency, and collocational transparency, syntactic complexity

achieved through phrasal compression would necessarily engender ambiguity and opaqueness for smooth comprehension (e.g. Biber & Gray, 2016; Elbow, 1991). Critical scholars such as Elbow (1991) warn of academic discourse as privileging students from the higher classes and excluding outsiders and its role in perpetuating the division between social classes. The downside of advocating linguistic complexity at the expense of meaningfulness and clarity is that students may learn to deploy the highly abstract and condensed features of academic discourse to produce what George Orwell (1946, p. 265) calls a language that is designed to "give an appearance of solidity to pure wind".

7.4. Recommendations for Future Research

Given the limitations of the study and renewed understandings in recent syntactic complexity research, it is necessary to recommend a number of new directions for future research to follow.

Continued efforts need to be made on the part of linguists to describe in-depth the highly flexible discourse realisations of grammatical devices to achieve syntactic complexity. Especially useful would be researching the interchangeability of highly synonymous grammatical choices (e.g. use of finite versus non-finite relative clauses) and the discourse-semantic conditions and authorial preferences that motivate the choices.

Another crucial area to be carefully explored would be how explicit language-focused instruction could enhance syntactic complexity. For advanced EFL students, instruction can focus on raising students' awareness of the functional and usage-based rationales for generating different dimensions of complex syntax. Experimental

research can be designed to assess the effects of explicit teaching on syntactic complexity growth within a time frame or in comparison with a control group.

What also deserves better understanding concerns the relationship between various dimensions of syntactic complexity and students' text quality for writing different genres. Although previous research has identified a positive correlation, careful qualitative investigation is needed of how specific syntactic arrangements contribute to coherence beyond mere correlation calculation. Quality-wise, future research should also take on board the issue of logical reasoning in syntactic complexity research, which by far has attracted little scholarly attention. Incorporating logicality is particularly crucial for studying the effectiveness of using finite adverbial clauses to build argumentation, considering the easy susceptibility of such logical relations as reason, concession, and condition to informal fallacy. Since attending to just the structural and formal at the expense of meaning does little to improve effective writing, future research on syntactic complexity is expected to divert attention to its usage-based and meaning aspects.

REFERENCES

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249-264). Amsterdam: John Benjamins.

Aitchison, C. (2009). Writing groups for doctoral education. *Studies in Higher Education, 34*(8), 905-916.

Allison, D., Cooley, L., Lewkowicz, J., & Nunan, D. (1998). Dissertation writing in action: The development of a dissertation writing support program for ESL graduate research students. *English for Specific Purposes, 17*(2), 199-217.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Baldwin, R. G., & Autin, A. E. (1995). Toward greater understanding of faculty research collaboration. *Review of Higher Education, 19*(1), 45-70.

Barton, E., Halter, E., McGee, N., & McNeilley, L. (1998). The awkward problem of awkward sentences. *Written Communication, 15*(1), 69-98.

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing, 22*(2), 185-200. doi:10.1007/s11145-007-9107-5

Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes, 43*(2), 79-120. doi:10.1207/s15326950dp4302_1

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243-257. doi:10.1093/llc/8.4.243

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

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English*. Essex: Pearson Education Limited.

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. (2016). *Grammatical complexity in academic English*. CUP.

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. (2014). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, Advance Access, 1-31. doi:10.1093/applin/amu059

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.

Bitchener, J., & Basturkmen, H. (2006). Perceptions of the difficulties of postgraduate L2 thesis students writing the discussion section. *Journal of English for Academic Purposes*, *5*(1), 4-18.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning, 60*(3), 534-572. doi:10.1111/j.1467-9922.2010.00566.x

Brumfit, C. J. (1984). *Communicative methodology in language teaching*. Cambridge University Press.

Bruner, J. (1986). *Actual minds, possible worlds*. New York: Plenum Press.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42-65. doi:10.1016/j.jslw.2014.09.005

Bunton, D. (1999). The use of higher level metatext in Ph. D theses. *English for Specific Purposes, 18*, S41-S56.

Bunton, D. (2002). Generic moves in PhD thesis introductions. In J. Flowerdew (Ed.), *Academic discourse* (pp. 57–75). London: Pearson Education.

Bunton, D. (2005). The structure of PhD conclusion chapters. *Journal of English for Academic Purposes, 4*(3), 207-224.

Burrough-Boenisch, J. (2003). Shapers of published NNS research articles. *Journal of Second Language Writing, 12*(3), 223-243. doi:10.1016/s1060-3743(03)00037-7

Byrnes, H. (2009). Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor. *Linguistics and Education, 20*(1), 50-66.

Byrnes, H., Maxim, H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment [Monograph]. *Modern Language Journal, 94*(S1).

Cadman, K. (1997). Thesis writing for international students: A question of identity? *English for Specific Purposes, 16*(1), 3-14.

Callies, M. (2013). Agentivity as a determinant of lexico-grammatical variation in L2 academic writing. *International Journal of Corpus Linguistics, 18*(3), 357-390. doi:10.1075/ijcl.18.3.05cal

Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35-53). New Jersey: ABLEX Publishing Corporation.

Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S.J. Samuels (Eds.), *Comprehending oral and written language* (pp. 83-113). New York: Academic Press.

Chafe, W. (1988). Punctuation and the prosody of written language. *Written Communication, 5*(4), 395-426.

Chan, A. Y. W. (2004). Syntactic transfer: Evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal, 88*(1), 56-74. doi:10.1111/j.0026-7902.2004.00218.x

Chang, C.F, & Kuo, C.H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes, 30* (3), 222-234.

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*(3), 310-331. doi:10.1016/j.esp.2005.05.003

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes, 6*(4), 289-302. doi:10.1016/j.jeap.2007.09.009

Chase, S. (1956). Preface. In J. B. Carroll (Ed.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (pp. v-x). Massachusetts: The MIT Press.

Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to concgram. *International journal of corpus linguistics, 11*(4), 411-433.

Cheng, X. & Wang, Q. (2004). Cong xiaoju guanxi kan xuesheng yingyu zuowen de lianguanxing [Looking at the coherence of EFL writing from the perspective of clause relations]. *Waiyu Jiaoxue Yu Yanjiu [Foreign Language Teaching and Research]*, *36*(4), 292-298.

CNKI. (2014). China Master's Theses Full-text Database. Retrieved February 02, 2016, from http://epub.cnki.net/kns/brief/result.aspx?dbPrefix=CDMD.

Coleman, J. A. (2006). English-medium teaching in European higher education. *Language Teaching, 39*(1), 1-14. doi:10.1017/s026144480600320x

Cooley, L. & Lewkowicz, J. (1997). Developing awareness of the rhetorical and linguistic conventions of writing a thesis in English: Addressing the needs of ESL/EFL postgraduate students. In A. Duszak (Ed.), *Culture and styles of academic discourse* (pp. 113-140). Berlin: Mouton de Gruyter.

Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research*, *69*(5), 176-183.

Cosme, C. (2008). Participle clauses in learner English: The role of transfer. In G. Gilquin, S. Papp, & M. B. Diez Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 177–200). New York, NY: Rodopi.

Council of Graduate Schools. (2013). Data *sources: International applications and admissions continue to grow*. Retrieved 18 May 2013 from http://www.cgsnet.org/data-sources-international-applications-and-admissions-continue-grow-0.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, *16*(3), 129-147. doi:10.1016/j.jslw.2007.07.002

Crammond, J. G. (1998). The uses and complexity of argument structures in expert and student persuasive writing. *Written Communication, 15*(2), 230-268.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing, 32*, 1-16. doi:10.1016/j.jslw.2016.01.003

Crowhurst, M. (1980a). Syntactic complexity in narration and argument at three grade levels. *Canadian Journal of Education / Revue canadienne de l'éducation, 5*(1), 6-12. doi:10.2307/1494634

Crowhurst, M. (1980b). Syntactic complexity and teachers' quality ratings of narrations and arguments. *Research in the Teaching of English, 14*(3), 223-231.

Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. *Language Proficiency and Academic Achievement, 10*, 2-19.

Elbow, P. (1991). Reflections on academic discourse: How it relates to freshmen and colleagues. *College English, 53*(2), 135. doi:10.2307/378193

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(02), 143-188.

Ellis, N. C., & Frey, E. (2009). The psycholinguistic reality of collocation and semantic prosody (2). In Corrigan, R., E. A. Moravcsik, H. Ouali, & K. M. Wheatley (Eds), *Formulaic language*: *Volume 2* (pp. 473-497). John Benjamins.

Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes, 6*(1), 3-17.

Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes, 30*(3), 198-208.

Evans, S., & Morrison, B. (2016). Adjusting to higher education in Hong Kong: the influence of school medium of instruction. *International Journal of Bilingual Education and Bilingualism*, 1-14. doi:10.1080/13670050.2016.1228600

Fairclough, N. (1989). *Language and power.* London: Longman.

Farsi, A. A. (1968). Classification of adjectives. *Language Learning, 18*(1-2), 45-60. doi:10.1111/j.1467-1770.1987.tb00389.x

Flowerdew, L. (2015). Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis. *Journal of English for Academic Purposes, 20*, 58-68. doi:10.1016/j.jeap.2015.06.001

Ford, C. E. (1994). Dialogic aspects of talk and writing: *Because* on the interactive-edited continuum. *Text*, *14*(4), 531-554.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305-327.

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41-61. doi:10.1075/etc.1.1.05gil

Gil-Salom, L., & Soler-Monreal, C. (2014). Writers' positioning in literature reviews in English and Spanish computing doctoral theses. *Journal of English for Academic Purposes, 16*, 23-39. doi:10.1016/j.jeap.2014.08.002

Givón, T. (2009). *The genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution*. John Benjamins Publishing.

Gosden, H. (1992). Research writing and NNSs: From the editors. *Journal of Second Language Writing, 1*(2), 123-139.

Grabe, W. & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. London, England: Longman

Granger, S. (1997). On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared. In J. Aarts, I. de Mönnink, & H. Wekker (Eds.) *Studies in English language and teaching* (pp. 185-198). Amsterdam & Atlanta: Rodopi.

Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.) *Learner English on computer* (pp. 119-131). London and New York: Longman.

Greaves, C. (2009). *ConcGram 1.0: A phraseological search engine* [Computer Software]. Amsterdam: John Benjamins.

Halliday, M. A. K. (1978). *Language as social semiotic*. Arnold: London.

Halliday, M. A. K. (1987). Spoken and written modes of meaning. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 55-82). New York: Academic Press.

Halliday, M. A. K. (1989). *Spoken and written language*. Oxford University Press.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education, 5*(2), 93-116.

Halliday, M.A.K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.

Halliday, M.A.K., & Martin, J.R. (Eds.). (1993). *Writing science: Literacy and discursive power*. Washington, D.C.: The Falmer Press.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (1999). *Construing experience through meaning: A language-based approach to cognition*. London: Continuum.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (2004). *An introduction to functional grammar*. London: Arnold.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. New Jersey: Lawrence Erlbaum Associates.

Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly, 37*(2), 275-301. Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473. doi:10.1093/applin/amp048

Hu, G. (2009). The craze for English-medium education in China: driving forces and looming consequences. *English Today, 25*(4), 47-54. doi:10.1017/s0266078409990472

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Research Report No. 3. Champaign, IL: National Council of Teachers of English.

Hyland, K. (1999). Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics, 20*(3), 341-367. doi:10.1093/applin/20.3.341

Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, *13*(2), 133-151.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London: Routledge.

Hyland, K. (2009). *Academic discourse: English in a global context*. London and New York: Continuum.

Hyland, K. (2016). Methods and methodologies in second language writing research. *System, 59*, 116-125. doi:10.1016/j.system.2016.05.002

Hyland, K., & Tse, P. (2005). Hooking the reader: a corpus study of evaluative *that* in abstracts. *English for Specific Purposes, 24*(2), 123-139. doi:10.1016/j.esp.2004.02.002

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly, 30*(4), 693-722.

IBM Corp. (2011). IBM SPSS Statistics for Windows (Version 20.0) [Computer Software]. Armonk, NY: IBM Corp.

Kang, H.-S. (2012). English-only instruction at Korean universities: Help or hindrance to higher learning? *English Today*, *28*(1), 29-34. doi:10.1017/s0266078411000654

Kaplan, R. B. (1966). Cultural thought patterns in inter-cultural education. *Language Learning*, *16*(1-2), 1-20.

Kawase, T. (2015). Metadiscourse in the introductions of PhD theses and research articles. *Journal of English for Academic Purposes, 20*, 114-124. doi:10.1016/j.jeap.2015.08.006

Kettemann, B., & Marko, G. (Eds.), *Teaching and learning by doing corpus analysis*. *Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz*. Amsterdam: Rodolphi.

Klein, D., & Manning, C. D. (2003). Fast exact inference with a factored model for natural language parsing. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 3-10). Cambridge, MA: MIT Press.

Kortmann, B. (1991). *Free adjuncts and absolutes in English*. London: Routledge.

Kwan, B. S. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, *25*(1), 30-55.

Kwan, B. S. (2008). The nexus of reading, writing and researching in the doctoral undertaking of humanities and social sciences: Implications for literature reviewing. *English for Specific Purposes, 27*(1), 42-56.

Kwan, B. S. (2009). Reading in preparation for writing a PhD thesis: Case studies of experiences. *Journal of English for Academic Purposes*, *8*(3), 180-191.

Larsen-Freeman, D. (2006). The Emergence of complexity, fluency, and accuracy in the oral and written Production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619. doi:10.1093/applin/aml029

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 579-589. doi:10.1093/applin/amp043

Larsen-Freeman, D. (2011). A complexity theory approach to second language development/acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 48-72). London: Routledge.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647-672. doi:10.1111/j.1467-9922.2010.00621.x

Lea, Diana. (2002). *Oxford collocations dictionary for students of English*. Oxford University Press.

Lee, J. J., & Casal, J. E. (2014). Metadiscourse in results and discussion chapters: A linguistic analysis of English and Spanish thesis writers in engineering. *System, 46*, 39-54. doi:10.1016/j.system.2014.07.009

Lee, S. C. N., & Tajino, A. (2008). Understanding Students' Perceptions of Difficulty with Academic Writing for Teacher Development: A Case Study of the University of Tokyo Writing Program. *京都大学高等教育研究*, *14*, 1-11.

Leki, I., Cumming, A., & Silva, T. (2010). *A synthesis of research on second language writing in English*. New York and London: Routledge.

Levy, R., & Andrew, G. (2006). Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*. Genoa, Italy: ELRA, 2231-2234.

Li, L.Y., & Vandermensbrugghe, J. (2011). Supporting the thesis writing process of international research students through an ongoing writing group. *Innovations in Education and Teaching International, 48*(2), 195-205.

Li, Y. (2006). A doctoral student of physics writing for publication: a socio-politically oriented case study. *English for Specific Purposes, 25,* 456-478.

Li, Y. (2007). Apprentice scholarly writing in a community of practice: An intraview of an NNES graduate student writing a research article. *TESOL Quarterly, 41*(1), 55-79.

Li, Y. (2015). "Publish SCI papers or no degree": practices of Chinese doctoral supervisors in response to the publication pressure on science students. *Asia Pacific Journal of Education, 36*(4), 545-558. doi:10.1080/02188791.2015.1005050

Li, Y., Hyland, F., & Hu, G. (2017). Prompting MEd students to engage with academia and the professional world through feedback. *Journal of English for Academic Purposes, 26,* 52-65. doi:10.1016/j.jeap.2017.02.005

Liang, M. (n. d.). Log-likelihood and chi-square calculator 1.0 (Version 1.0) [Computer Software]. Beijing, China: Beijing Foreign Studies University.

Lim, J. M.-H., Loi, C.-K., Hashim, A., & Liu, M. S.-M. (2015). Purpose statements in experimental doctoral dissertations submitted to U.S. universities: An inquiry

into doctoral students' communicative resources in language education. *Journal of English for Academic Purposes, 20*, 69-89. doi:10.1016/j.jeap.2015.06.002

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes, 31*(3), 150-160. doi:10.1016/j.esp.2011.10.002

Liu, L., & Li, L. (2016). Noun phrase complexity in EFL academic writing: A corpus-based study of postgraduate academic writing. *The Journal of Asia TEFL, 13*(1), 48-65. doi:10.18823/asiatefl.2016.13.1.4.48.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, D. Coste, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: Benjamins.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474-496.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*(1), 36-62.

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

Lu, X. (2014). *Computational methods for corpus annotation and analysis*. Springer.

Lu, X. (n.d.). Lexical Complexity Analyzer. Retrieved August 18, 2015, from http://www.personal.psu.edu/xxl13/downloads/lca.html.

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing, 29*, 16-27. doi:10.1016/j.jslw.2015.06.003

MacDonald, S. P. (1992). A method for analyzing sentence-level differences in disciplinary knowledge making. *Written Communication*, *9*(4), 533-569.

Madsen, D. (1992). *Successful dissertations and theses: A guide to graduate student research from proposal to completion* (2nd ed.). San Francisco: Jossey-Bass.

Manalo, E. (2006). The usefulness of an intensive preparatory course for EAL thesis writers. *Journal of Research in International Education*, *5*(2), 215-230.

Martin, J. R. (1992). English text: *System and structure*. Amsterdam: John Benjamins.

Martin, J. R. (1993). Life as a noun: Arresting the universe in science and humanities. In M. A. K. Halliday, & J.R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 221-267). University of Pittsburgh Press

Martin, J. R., & White, P. R. (2005). *The language of evaluation*. London: Palgrave Macmillan.

Master, P. (1991). Active verbs with inanimate subjects in scientific prose. *English for Specific Purposes, 10*(1), 15-33. doi:10.1016/0889-4906(91)90013-m

Maton, K., & Muller, J. (2007) A sociology for the transmission of knowledges, in F. Christie & J. Martin (Eds.), *Language, knowledge and pedagogy* (pp. 14-33). London: Continuum.

Matthiessen, C. M. I. M., Teruya, K., & Lam, M. (2010). *Key terms in systemic functional linguistics*. New York/London: Continuum.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2009). Linguistic features of writing quality. *Written Communication, 27*(1), 57-86. doi:10.1177/0741088309351547

Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge University Press.

Mohan, B. A., & Lo, W. A.-Y. (1985). Academic writing and Chinese students: Transfer and developmental factors. *TESOL Quarterly, 19*(3), 515. doi:10.2307/3586276

Neff, J., Dafouz, E., Diez, M., & Prieto, R. (1998). Contrastive discourse analysis: Argumentative text in English and Spanish. Paper presented at the Twenty-Fourth Linguistics Symposium: Discourse across Languages and Cultures, University of Wisconsin-Milwaukee, September.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in SLA: The case of complexity. *Applied Linguistics, 30*, 555-578.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223-242. doi:10.1093/applin/24.2.223

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

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing, 29*, 82-94. doi:10.1016/j.jslw.2015.06.008

Orwell, G. (1946) Politics and the English language. *Horizon, 13* (April), 252-265.

Paltridge, B. (1997). Thesis and dissertation writing: Preparing ESL students for research. *English for Specific Purposes, 16*(1), 61-70. doi:10.1016/s0889-4906(96)00028-2

Paltridge, B. (2002) Thesis and dissertation writing: An examination of published advice and actual practice. *English for Specific Purposes, 21*, 125-143.

Paltridge, B., & Starfield, S. (2007). *Thesis and dissertation writing in a second language: A handbook for supervisors*. Routledge.

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. London and New York: Continnum.

Parkinson, J. (2015). Noun–noun collocations in learner writing. *Journal of English for Academic Purposes, 20*, 103-113. doi:10.1016/j.jeap.2015.08.003

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.

Plaks, A. H. (1988). Where the lines meet: Parallelism in Chinese and Western literatures. *Chinese Literature: Essays, Articles, Reviews (CLEAR), 10*(1/2), 43. doi:10.2307/495141

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning, 64*, 878-912.

Polio, C. (2001). Research methodology in second language writing research: The case of text-based studies. In T. Silva & P. K. Matsuda (Eds.), *On second language writing* (pp. 91–115). Mahwah, New Jersey: Lawrence Erlbaum.

Qian, J., & Krugly-Smolska, E. (2008). Chinese graduate students' experiences with writing a literature review. *TESL Canada Journal, 26*(1), 68-86.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Raimes, A. (1990). The TOEFL Test of Written English: Causes for Concern. *TESOL Quarterly, 24*(3), 427-442. doi:Doi 10.2307/3587228

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Richards, R. T. (1988). Thesis/dissertation writing for EFL students: An ESP course design. *English for Specific Purposes, 7*(3), 171-180.

Ryshina-Pankova, M. (2015). A meaning-based approach to the study of complexity in L2 writing: The case of grammatical metaphor. *Journal of Second Language Writing, 29*, 51-63. doi:10.1016/j.jslw.2015.06.005

San Miguel, C., & Nelson, C. D. (2007). Key writing challenges of practice-based doctorates. *Journal of English for Academic Purposes, 6*(1), 71-86.

Sawaki, T. (2014). On the function of stance-neutral formulations: Apparent neutrality as a powerful stance constructing resource. *Journal of English for Academic Purposes, 16*, 81-92. doi:10.1016/j.jeap.2014.10.001

Schachter, J. (1974). An error in error analysis. *Language Learning*, *24*(2), 205-214.

Schleppegrell, M. J. (1996). Conjunction in spoken English and ESL writing. *Applied Linguistics, 17*(3), 271-285.

Schleppegrell, M. J. (1998). Grammar as resource: Writing a description. *Research in the Teaching of English, 32*(2), 182-211.

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education, 12*(4), 431-459.

Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Erlbaum.

Scott, M. (2008). WordSmith Tools (Version 5.0) [Computer Software]. Liverpool, UK: Lexical Analysis Software.

Shaw, P. (1991). Science research students' composing processes. *English for Specific Purposes, 10*(3), 189-206. doi:10.1016/0889-4906(91)90024-q

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly, 27*(4), 657. doi:10.2307/3587400

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Skehan, P. (1989). *Individual differences in second language learning*. Edward Arnold.

Smith, C. S. (2003). *Modes of discourse: The local structure of texts*. Cambridge University Press.

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. doi:10.1177/0741088316631527

Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity. *Written Communication*, *18*(4), 506-548. doi: 10.1177/0741088301018004004

Stapleton, P., & Wu, Y. (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes, 17*, 12-23. doi:10.1016/j.jeap.2014.11.006

Starfield, S., & Ravelli, L. (2006). The writing of this thesis was a process that I could not explore with the positivistic detachment of the classical sociologist: Self and structure in New Humanities research theses. *Journal of English for Academic Purposes, 5*, 222-243.

Swales, J. (1981). *Aspects of article introductions*. Language Studies Unit, University of Aston in Birmingham.

Swales, J. (1990). *Genre analysis*: *English in academic and research settings*. Cambridge University Press.

Swales, J. (2004). *Research genres: explorations and applications*. Cambridge University Press.

Tannen, D. (1982). *Spoken and written language: Exploring orality and literacy*. New Jersey: ABLEX Publishing Corporation.

Tedick, D. J. (1988). *The effects of topic familiarity on the writing performance of non-native writers of English at the graduate level.* Unpublished doctoral dissertation, Ohio State University, the United States.

The State Council. (2015). *Report on the work of the government delivered at the third session of the 12th National People's Congress on March 5, 2015*. Retrieved 10 March from http://english.gov.cn/archive/publications/2015/03/05/content_281475066179954.htm.

Thomas, S., & Hawes, T. P. (1994). Reporting verbs in medical journal articles. *English for Specific Purposes, 13*(2), 129-148. doi:10.1016/0889-4906(94)90012-4

Thompson, G., & Yiyun, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics*, *12*(4), 365-382. doi:10.1093/applin/12.4.365

Thompson, P. (2001). *A pedagogically-motivated corpus-based examination of PhD theses: Macrostructure, citation practices and uses of modal verbs.* Unpublished doctoral dissertation, University of Reading, the United Kingdom.

Thompson, P. (2012). Thesis and dissertation writing. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 283-299). Chichester: Wiley-Blackwell.

Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning and Technology*, *5*(3), 91-105.

Thompson, S. A. (1991). On addressing functional explanation in linguistics. *Language & Communication*, *11*(1), 93-96.

Toulmin, S. E. (1958). *The uses of argument.* Cambridge University Press.

Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003*, pp. 252-259.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, *40*(1), 21-81.

Vande Kopple, W. J. (1994). Some characteristics and functions of grammatical subjects in scientific discourse. *Written Communication*, *11*(4), 534-564.

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal*, *96*(4), 576-598.

Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, *97*(S1), 1-20.

Warren, M. (2016). Data-driven learning. In F. Farr & L. Murray (eds.) *Routledge handbook of language learning and technology* (pp. 61-79). London: Routledge.

Wen, Q., Wang, L., & Liang, M. (2005). *Spoken and written English corpus of Chinese learners*. Beijing: Foreign Language Teaching and Research Press.

Williams, E. (1978). Across-the-board rule application. *Linguistic Inquiry*, *9*(1), 31-43.

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

Xie, J. (2016). Direct or indirect? Critical or uncritical? Evaluation in Chinese English-major MA thesis literature reviews. *Journal of English for Academic Purposes, 23*, 1-15. doi:10.1016/j.jeap.2016.05.001

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.

Yau, M. S. S., & Belanger, J. (1984). The influence mode on the syntactic complexity of EFL students at three grade levels. *TESL Canada journal, 2*(1), 65. doi:10.18806/tesl.v2i1.450

Yeh, C. C. (2010). New graduate students' perspectives on research writing in English: A case study in Taiwan. *Journal of Academic Language and Learning, 4*(1), A1-A12.

Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes, 10*(3), 130-139.

Zhang, Y. (2014). *Adversative and concessive conjunctions in Chinese EFL writing: A functional perspective*. Unpublished doctoral thesis, The Hong Kong Polytechnic University, Hong Kong SAR, China PR.