Genre Features of Personal Statements by Chinese English-as-an-Additional-Language Writers: A Corpus-Driven Study

by

Sibo Chen

B.A., Hebei University, 2011

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

Supervisory Committee

Genre Features of Personal Statements by Chinese English-as-an-Additional-Language Writers: A Corpus-Driven Study

by

Sibo Chen

B.A., Hebei University, 2011

Supervisory Committee

Dr. Li-Shih Huang (Department of Linguistics)
Supervisor
Dr. Hossein Nassaji (Department of Linguistics)
Departmental Member

Abstract

Supervisory Committee
Dr. Li-Shih Huang (Department of Linguistics)
Supervisor
Dr. Hossein Nassaji (Department of Linguistics)
Departmental Member

Personal Statements (PSs) are self-narrative essays written for Western graduate school applications, which serve an important role in Western graduate schools' admission processes. However, genre features of PSs have not been sufficiently addressed by previous genre studies. Such neglect indicates a promising area for investigation as the increasing number of non-native English speakers in Western higher education systems creates an urgent pedagogical need for PS-related English-as-an-Additional-Language (EAL) instruction.

The present thesis reports a corpus-driven genre analysis of PSs written by Chinese EAL students (CEAL-PSs). Based on a corpus of 120 CEAL-PS samples, genre features of CEAL-PSs were investigated from three perspectives: (1) linguistic complexity (i.e. lexical diversity and grammatical intricacy), (2) content foci (i.e. at the lexical, phrasal, discoursal levels), and (3) functional move structure. In addition, comparative analyses were made between unedited and edited CEAL-PSs for investigating whether the editing process significantly changed the unedited CEAL-PSs in the above three perspectives.

There were three major findings of the current study. First, the majority of lexicons used by the collected CEAL-PSs were frequent academic lexicons and the average grammatical intricacy of these samples was at senior high school or junior college levels. Second, expressions of self-promotion and discussions of academic/professional achievements were explicitly emphasized in the collected CEAL-PSs at the lexical, phrasal, and discoursal levels. Third, an IERC model ("Introduction," "Establishing Credentials," "Reasons for Application," and "Conclusion"), was found to be followed by the majority of the collected CEAL-PSs. Based on the above findings, the thesis further discusses the current study's theoretical, methodological, and pedagogical implications for EAL writing instruction in China.

Table of Contents

List of Tables

List of Figures

Acknowledgments

As academic folklore says, a thesis can never be a solitary project. I would like to give my special thanks to the following people for their enormous help during the completion of the project.

First, I would like to give my special thanks to my supervisor committee, Dr. Li-Shih Huang and Dr. Hossein Nassaji for their diligent reviews and constructive comments. In particular, they offered me invaluable support when I was away from the University of Victoria. The preliminary idea of this thesis came from my first graduate course at UVic, which was taught by Dr. Huang. Along with my progress at UVic, Dr. Huang has not only provided enormous guidance for my thesis research, she also inspired my desire for pursuing PHD research by her dedication to academia. I first met Dr. Nassaji at AILA 2011 in Beijing before my graduate studies at UVic. Since then, Dr. Nassaji has become a great mentor for me. He has offered constructive comments during the development of the thesis and valuable comments during the versions of the manuscript.

Furthermore, I would like to thank Dr. Erin Kelly for her thorogh review of the thesis. The thesis was defended during Dr. Kelly's research trip, which cannot be done without her kindly accommodation. I also want to thank all of my friends at the Department of Linguistics, University of Victoria, who have provided invaluable encouragement for my academic adventure in Canada.

In addition, I would like to thank my audience at Genre 2012 and AACL 2013, especially Dr. Vijay Bhatia and Dr. Mike Scott for their constructive feedback for presentations of the thesis. Dr. Bhatia provided me with his insights of promotional genre, which laid the theoretical foundation of the present study. Dr. Scott, with his fabulous software WordSmith Tools, made my life much easier during data analysis.

Finally, I want to give my special thanks to my parents. It is no need to say that without their support, my adventure in Canada can never be realized.

Despite several rounds of editing, the present manuscript may still bear some typos and inappropriate expressions. All errors are my own.

Dedication

To my parents, who support my adventure in the unpredictable world

Chapter 1 Introduction

Genres are defined as particular forms of discourse with shared structure, style, content, and intended audience (Swales, 1990). Recent publications in applied linguistics have defined genres as "sociolinguistic activities used by genre producers to achieve particular communicative goals" (Henry & Roseberry, 2001, p. 153).

As one key area in applied linguistics, genre research has significant theoretical and pedagogical implications. Theoretically, genre studies are able to explore different types of texts' linguistic features and analyse genre usage in daily communications from both macro and micro perspectives. Pedagogically, genre studies' findings can offer valuable suggestions for writing instruction, especially for English-as-an-Additional-Language (EAL) writing instruction.

Although a variety of academic genres (e.g. abstracts, research presentations, and grant proposals) has been extensively investigated by previous studies (e.g. Bhatia, 1993; Hyland, 2009; Samraj, 2002, 2005; Swales, 1990), one academic genre, personal statements, has not been sufficiently addressed by previous genre literature. Personal statements (PSs) (also known as letters of application, statements of research interests, or statements of purposes) are self-narrative essays written for Western graduate school applications. In a typical PS, an author usually describes his/her education background, research interests, personalities and relevant qualifications in order to pursue acceptance of his/her desired graduate program. According to previous studies (Brown 2004; Ding, 2007; Samraj & Monk, 2008), PSs can be defined as an academic self-promotional genre. However, such definition has not been completely confirmed as studies on PSs have been very limited in the past decade.

Compared with other academic genres, PSs have several unique features, which create great challenges for many graduate school applicants who do not have sufficient PS writing experience. First, compared with academic genres with relatively rigid structure requirements (e.g. journal articles, research reports, and grant proposals), PSs allow a certain degree of creativity, which causes considerable variations within this genre (Brown, 2004).

Second, different academic disciplines may have different or even conflicting preferences of PS content and organization. For example, descriptions of personal experience are regarded as a peripheral component in PSs written for clinical psychology programs whereas they are an indispensable section for PSs written for medical/dental schools (Brown, 2004; Ding, 2007).

Third, the PS genre is rarely, if ever, taught in university writing classes. Institutional support for PS writing, such as workshops offered by university writing centres and libraries, is also scarce. The informal survey in Brown (2004) showed that consultations of PS writing constituted a sizable proportion of inquiries received by one major U.S. university writing centre whereas these inquiries can rarely be satisfactorily solved due to the writing centre advisors' lack of relevant genre knowledge. Meanwhile, commercial guidebooks of PS writing can be ineffective or even misleading. The reviews of these guidebooks in previous studies have shown that this type of guidebooks often aims at reaching the largest audience group for sale purposes (Brown, 2004; Samraj & Monk, 2008). Consequently, most guidelines provided by these books are often very generic without any disciplinary specification. These guidelines are also non-empirical and only based on personal experience offered by "admission experts."

Fourth, graduate schools' admission websites usually provide insufficient support for prospective students in terms of PS writing. Graduate schools' admission websites are regarded as the most reliable online resources by their applicants; however, information regarding PS content/format requirements is often missing on these websites. Before writing this thesis, I conducted an informal survey on "Graduate Admission" pages of 30 departments in several major Canadian universities. It turned out that only 11 departments described some specific requirements for PS writing, and among them, only four offered explicit requirements (e.g. preferred PS structure and content) for their applicants.

In sum, despite graduate school applicants' urgent need for PS writing advice, very few studies have explored the genre's content and structures. In particular, the scarcity of research on the PS genre is disappointing. In the past decade, only three articles on the PS genre have appeared in major applied linguistics journals (i.e. Brown, 2004; Ding, 2007; Samraj & Monk, 2008) and all of them have focused on the research of the PS genre's textual structure (i.e. functional moves). Consequently, other aspects of the PS genre, such as linguistic complexity and content foci, have not been sufficiently explored by previous genre literature. Furthermore, in recent years online essay editing services have become a prominent phenomenon. However, it is unknown whether PSs edited by these self-claimed professionals can be significantly different from the unedited ones. Thus, the current situation of the PS genre proposes a valuable research topic, which requires an investigation from multiple perspectives.

The current thesis reports a corpus-driven study of the PS genre, focusing on PSs written by Chinese EAL applicants (hereafter referred as CEAL-PSs). Based on a total of

120 PS samples (a total of 112,840 words/tokens), the present study analysed the collected CEAL-PSs' linguistic complexity, content foci, and functional move structure. Comparative analyses were also conducted between edited and unedited CEAL-PSs for investigating whether the editing process significantly changed the unedited CEAL-PSs in the above three perspectives.

The thesis is organized as follows. Chapter 2 reviews the history of genre research and previous studies of the PS genre. Chapter 3 introduces the methodology of the current study, in which the data collection and analysis procedure are described. Chapter 4 provides the results of data analysis. Finally, Chapter 5 discusses of the current study's major findings and theoretical, methodological, and pedagogical implications.

Chapter 2 Literature Review

2.1 Defining Personal Statement from a Genre Perspective

The English term "genre" derives from the French lexicon "genre," which originally comes from the Latin word "genus" with the meanings of "type" and "kind" (Corbett, 2006). In academia, the term "genre" is traditionally used in literacy, film, and music studies, referring to any category of works that are recognizable by their adherence to conventions of form, content, and language style (Corbett, 2006). In the field of discourse analysis, the concept of "genre" has been widely recognized since the 1950s. One early definition of "genre" came from Bakhtin (1953), in which "genre" was defined as a notion at the discoursal level, referring to "types of utterances that are heterogeneous in nature" (p. 60). Bakhtin further argued that texts sharing similar forms, content, and language styles can be recognized as one type of genre.

Following Bakhtin, the notion of "genre" has been expanded by three schools since the 1980s: the Systemic Functional Grammar (SFG) School, the New Rhetoric (NR) School, and the English for Specific Purposes (ESP) School (Hyon, 1996). The SFG School emphasizes that texts belonging to one specific genre are inseparable from their communicative purposes and target audience. For instance, Halliday and Hasan (1989) argued that a genre can be identified by the obligatory elements in its textual structure and attached contextual configurations. By comparison, the NR School pays more attention to the users (both producers and audience) of different genres. Miller (1984) described genres as social actions, focusing "not on the substance or the form of discourse, but on the communicative purposes it is used to accomplish" (Miller, 1984, p. 151). The ESP School is probably the most influential school in genre studies. Derived

from a pedagogical perspective, professionals in the ESP School have suggested that the findings of textual variations of different genres can be effectively applied in ESP writing instructions. As a result, genre studies conducted by the ESP School have tended to investigate structural variations of different genres and their potential pedagogical implications for academic and professional purposes (e.g. Bhatia, 1993; Swales, 1990).

There are two essential concepts in contemporary genre studies: "discourse community" and "move." "Discourse community" is defined as a community that shares a common set of communicative purposes (Swales, 1990). It derives from the "speech community" concept in sociolinguistics, which refers to a community of members who share similar language rules (Saville-Troike, 1982). For instance, people sharing the same vernacular in New York City, as studied in Labov (1966), can be recognized as belonging to the same speech community. Similarly, in genre studies "discourse community" is defined as a community that shares the same genre as the primary channel of communication. Swales (1990) proposed six essential components for a discourse community (Table 1).

Table 1 ***Features of a Discourse Community***

A discourse community has:
1) A broadly agreed set of common public goals
2) Mechanisms of inter-communication among its members
3) Its participatory mechanisms primarily to provide information and feedback
4) One or more genres in the communicative furtherance of its aims
5) Some specific lexis
6) A threshold level of members with a suitable degree of relevant content and discoursal expertise

Adapted from Swales (1990, pp. 24-27)

To be specific, in Table 1, features (1) to (3) construct the socio-cultural foundations of different types of genres. By contrast, the following features (4) to (6) contribute to the emergence of a specific genre among the members of one discourse community. With the introduction of "discourse community," different genres are not only rules regulating

linguistic forms, but also tools constructing social relationships. The textual structures of genres belonging to the same discourse community are significantly influenced by the communicative purposes and characteristics of members in that discourse community (Swales, 1990).

The concept of "move" originates from the studies conducted by the ESP School. Swales (1990) defined "move" as a functional unit of texts used for identifiable purposes and subsequently regulating the structures and content of a specific genre. Texts belonging to one genre are expected to show similar move structures. For instance, Zappen (1983) identified a problem-solution model[1] for the introduction section of academic journal articles, which, according to Zappen, can be found in many research articles written for science/social science journals.

Furthermore, by following the move structures of various genres, texts belonging to the same genre can be more easily recognized by the members of that genre's target discourse community. In other words, move structures make significant contributions to the fulfilment of genres' communicative purposes. Swales (1990) proposed the common structure in research articles (i.e. abstract, introduction, literature review, methodology, results, discussion, and conclusion) as a result driven by the communicative purposes of academic communities. Similarly, texts belonging to the same professional genre often have identical content and structures due to the influence of underlying move regulations of their target discourse communities (Bhatia, 1993). Therefore, studying different genres' move structures can offer valuable insights of genre production, as well as suggestions for the learners of different genres.

[1] There are five moves in the problem-solution model: Move 1-Goal, Move 2-Current Capacity, Move 3-Problem, Move 4-Solution, Move 5-Criteria of evaluation.

In sum, genres are various types of textual constructions that combine particular linguistic features and sociocultural contexts. For genre producers, the primary purpose of using a genre is to realize socialization in the genre's target discourse communities. The basic structural unit of genres is move. Following the above definitions, the next section will define PSs as an academic genre with occluded and promotional features.

2.2 Personal Statement as an Academic Genre

Due to the lack of relevant studies, PSs have not been convincingly defined from a genre perspective so far. In this section, I will define the PS genre from three perspectives. To be specific, PSs are: (1) an academic genre produced by graduate school applicants, especially senior level undergraduates, (2) an occluded genre with limited access to novice PS writers, and (3) a self-promotional genre resembling many professional genres (e.g. cover letters and grant proposals).

Genre research in the early development stage was closely associated with the fields of EAP (English for Academic Purposes) and ESP (English for Specific Purposes) (e.g. Dudley-Evans, 1994; Hopkins & Dudley-Evans, 1988; Morrow, 1989; Salager-Meyer, 1990; Salager-Meyer, Defives, Jensen, & De Filipis, 1989; Swales, 1981, 1990). One major purpose of early EAP/ESP studies was to explore English usage in different contexts thereby satisfying the pedagogical needs of various EAL learners. As a result, early genre studies tended to focus on texts belonging to genres with rigid structures, such as research articles, medical reports, and business letters. For instance, Swales (1981, 1990) studied different sections of research articles and established general move structures to guide the writing of research articles. Salager-Meyer analysed the discoursal flaws in medical English abstracts (e.g. the lack of essential moves or illogical orders of moves) in a series of studies (Salager-Meyer, 1989, 1990; Salager-Meyer et al., 1989).

Professional genres were also extensively addressed by early genre studies. Bhatia (1993) analysed a variety of professional texts, such as advertisements, sales promotional letters, and cover letters for job applications. Dudley-Evans (1994) and Flowerdew (1990) offered comprehensive reviews of professional genre studies.

By comparison, student writings (e.g. course assignments, research reports, and term papers) were rarely investigated by early genre studies. One reason for such neglect was the impression of "immaturity" in student writings due to students' undeveloped writing skills and genre knowledge. Compared with formal texts (e.g. academic journal articles), student writings used to be considered as texts with grammatical errors and irregular variations, which would cause extra difficulties for genre analysis. In recent years, although many valuable studies have been conducted on texts by students at senior levels (e.g. Hyland 2003, 2009; Swales, 1990), less formal writings by undergraduates (e.g. course assignments and term papers) have only been addressed by a few publications (Hyland, 2009; Samraj & Monk, 2008). Thus, as one less formal student genre, PSs have been "virtually ignored by previous genre studies" (Brown, 2004, p. 242).

The second facet of the PS genre is its occluded feature, which obstructs novice PS writers' full access to previous PS samples. The notion of "occluded genre" is defined as genres without open access to the public (Swales, 1996). Many genres cannot be easily accessed by people outside particular discourse communities due to confidential regulations, which create additional obstructions for people who want to learn these genres.

In Swales (1996), submission letters for academic journals were studied. Swales argued that, although texts such as submission letters and email communications between editors

and authors facilitated the production of research articles, these texts were not shown in the final manuscripts and thus occluded from academic journal readers. Swales (1996) further provided a list of typical occluded genres (e.g. retention-promotion tenure reports, peer referee reports, and business school application essays), which were further investigated by several follow-up studies (e.g. Fortanet, 2008; Hyon, 2008; Loudermilk, 2007). One common finding of these studies was that occluded genres cause considerable difficulties for novice writers.

Despite the PS genre's absence in Swales' occluded genre list, it can be defined as an occluded genre for the following reasons. First, access to PS samples already submitted to graduate schools is often restricted to admission committees due to privacy regulations. PSs are collected annually by academic departments for admission purposes and then destroyed after a certain preservation period, which limits their availability to novice PS writers. Second, many applicants may be reluctant to share their PSs with others due to privacy concerns. Third, suggestions for PS writing offered by guidebooks of graduate school applications may not be accurate. As discussed in Chapter 1, the suggestions offered by guidebooks are too general to effectively assist applicants who need specific suggestions for different academic disciplines. Fourth, resources for PS writing from university writing centres are limited due to writing consultants' unfamiliarity with the PS genre. In sum, the above factors make PSs an occluded genre.

The third facet of the PS genre is its self-promotional purpose. Promotional genres are genres with promotion as their primary communicative purpose (Bhatia, 1993). Bhatia (1993) reviewed common genres in professional contexts (e.g. job application letters, advertisements, and sales promotion letters) and categorized them as "promotional genres"

due to their similar communicative purposes, modes, and structures (Bhatia, 1993, p. 59). Bhatia (1993) further identified five typical features of job application letters (Table 2), which can also be observed in many PS samples.

Table 2 ***Typical Features of Job Application Letters***

1.	Job application letters are used for a persuasive purpose, which is often realized via eliciting a positive response of readers.
2.	Credentials establishment is the most pertinent opening of many job application letters.
3.	In job application letters written in response to advertisements, creating relevant descriptions for job requirements is often found.
4.	Job application letters often only contain essential information for the sake of precision.
5.	From an interpersonal perspective, job application letters are used to initiate positive relations between the applicants and the employers.

Adapted from Bhatia (1993, pp. 59-60)

For instance, the primary communicative purpose of PSs is to facilitate graduate school applications. In PSs, students often emphasize their academic/professional achievements to show their qualifications for graduate program admission requirements. Politeness is considered by many PS writers due to the power disparity between the admission committees and the applicants. Many PS writers prefer discussing their achievements in a humble way to create positive impressions for admission committee members (Ding, 2007).

In sum, the PS genre is a complex academic genre with occluded and self-promotional features, which can partially account for the lack of relevant studies on the PS genre. The next section will critically review three previous studies of the PS genre and discuss their key findings and limitations.

2.3 Previous Studies on the PS Genre

Previous studies of the PS genre are very rare. Before conducting the present study, I conducted a thorough literature search in several linguistics databases (e.g. Web of Sciences, ProQuest, and Sciences Direct) with key words such as “personal statement”

and "genre" and found that relevant studies on the PS genre in the past decade only included three journal articles published by *Written Communication* and *English for Specific Purposes* (Brown, 2004; Ding, 2007; Samraj & Monk, 2008), one special issue of *Issues in Writing*, and several unpublished degree theses from English or Education departments. This section will critically review the three studies published by published by *Written Communication* and *English for Specific Purposes* and discuss their key findings and limitations.

Brown (2004) is the first study of the PS genre appearing in a major applied linguistics journal (*Written Communication*). In this study, Brown analysed T-units in PSs written for a clinical psychology program. "T-unit" refers to an independent clause with embedded clauses and modifying phrases (Hunt, 1965, as cited in Brown, 2004, p. 248). The analyses in Brown (2004) showed that: (1) for clinical psychology programs, PSs by successful applicants contained more T-units describing research experience and interests than those by unsuccessful applicants; (2) one crucial task for PS writers was to prove their familiarization of their target discourse communities (clinical psychology in Brown's study); (3) interviews with admission committee members revealed three essential aspects of scientific discourse expected in PS texts (Table 3).

Table 3 Essential Aspects in Successful PSs for Clinical Psychology Programs

1.	Study Intuition to Scientific Empiricism: how does an applicant transfer his/her intuitive interest of psychology to empirical inquiry of it?
2.	Practical Application to Research Orientation: how does an applicant describe his/her research goals from a theoretical perspective?
3.	Egocentrism to Communitarianism: how does an applicant express his/her career goals, not only for self-actualization, but also for the sake of public interests?

Adapted from Brown (2004, p. 252)

Following the findings of Brown (2004), Ding (2007) studied the functional move structure of PSs written for medical/dental school applications. There were three key

findings in Ding (2007). First, a framework of move structure of the PS genre was established (Table 4), in which the first four moves (i.e. Reason for Application, Establishing Credentials, Relevant Life Experience, and Future Career Goals) were defined as obligatory moves whereas the last move (Personality) was defined as an optional move.

Table 4 ***The PS Move Structure in Ding (2007)***

Move 1: Reasons for Application
• Step 1: academic/intellectual interests
• Step 2: understanding of the field
• Step 3: personal/family experiences
Move 2: Establishing Credentials
• Step 1: academic achievements
• Step 2: research experiences
• Step 3: professional experiences
Move 3: Relevant Life Experiences
Move 4: Future Career Goals
Move 5: Personality

Adapted from Ding (2007, p. 378)

Second, Ding's analysis of move frequencies revealed that successful PSs for medical/dental schools devoted more effort to develop the obligatory moves, especially move 1 (Reason for Application) and move 2 (Establishing Credentials). By comparison, unsuccessful PSs spent more paragraphs describing the applicants' good characters.

Third, there were fewer personal stories used in the successful PS samples compared with unsuccessful ones. In particular, the stories used by successful PS samples were closely associated with moves 1 and 2. Overall, Ding's study confirmed previous findings in Brown (2004) and suggested that the discussion of research interests was expected to be the major content in successful PSs.

Samraj and Monk (2008) is the most recent study of the PS genre. Similar to Brown (2004), this study included a move analysis of PS samples and an interview with graduate

program chairs. Compared with the research designs of previous two studies, Samraj and Monk (2008) had two modifications. First, the study included a survey of books and websites regarding PS writing, which investigated the current resources available for novice PS writers. Second, the study collected PS samples from three disciplines (i.e. MBA, Electronic Engineering, and Linguistics) instead of one, which generated an inquiry of potential cross-disciplinary structural variations of the PS genre.

There were three key findings in Samraj and Monk (2008). First, the survey of PS-related books and websites showed that although suggestions for PS writing were not difficult to find online, these suggestions had two major limitations: (1) professional graduate programs such as MBA and MD (Doctor of Medicine) were better covered than research oriented programs, and (2) no clear distinction was made between PS writing suggestions offered for master and doctoral programs.

Second, based on a move analysis of 35 PS samples of three disciplines, Samraj and Monk revised the functional move structure proposed in Ding (2007) (Table 5).

Table 5 ***The PS Move Structure in Samraj and Monk (2008)***

1. Background
• General
• Work
• Education
• Research
• Personal Attributes
2. Reasons for Applying
• Gap in Background
• Positive Gains
• Program/University Attributes
• Disciplinary/Research Reasons
3. Extra-curricular Information
4. Conclusion
• Goals/Prediction of Future
• Self-evaluation

Adapted from Samraj and Monk (2008, p. 201)

The cross-disciplinary inquiry in Samraj and Monk (2008) revealed that the discussion of research interests was not always the content focus in PS writing. The data analysis of Samraj and Monk (2008) indicated that the discussion of research interests was only emphasized by PSs written for the electronic engineering program. By contrast, PSs written for the MBA and linguistics programs tended to emphasize the gaps between the applicants' current knowledge and their future academic/career goals. Third, the interviews with graduate program chairs revealed the conflicting views of the occluded status of the PS genre among graduate admission committees. The interviewed graduate program chairs admitted the difficulties students encountered during PS writing; however, they also expressed concern of the sincerity of PSs if detailed information was provided to applicants. In sum, Samraj and Monk's study highlighted the disciplinary variations of the PS genre and showed that discussions of research interests may not be unanimously focused by PSs in all disciplines.

Although the reviewed studies have offered many interesting findings of the PS genre, they are limited in three perspectives. First, all the reviewed studies included very limited data. In each study, only 30 to 35 PSs were analysed and these samples only came from one to three disciplines. Such inadequacy of data inevitably compromised the reliability and generalizability of the reviewed studies. Although Samraj and Monk (2008) attempted to solve the problem by analysing PS samples from three disciplines, PSs written for natural science and social science programs, such as physics, biology and sociology have not been analysed so far. Second, the reviewed studies mainly used PSs written by native English speakers as their research data. Consequently, genre features of PSs produced by EAL writers have not been sufficiently explored. Although Samraj and

Monk (2008) briefly addressed the issue of EAL writing in PS production, the study's limited sample size did not generate any concrete findings. The investigation of PSs by EAL writers is an extremely important issue since Western universities have witnessed a dramatic increase of international students in recent years. Third, all the reviewed studies adopted hand tagging methods (i.e. manually code text and conduct qualitative analysis) for their data analyses, which mainly focused on the macro features (e.g. textual construction) of the PS genre. By contrast, micro features of the PS genre, such as its lexico-grammatical features, have been neglected.

To address the reviewed studies' limitations, the current study adopted a corpus-driven research design to investigate PSs by Chinese EAL applicants (CEAL-PSs) from three perspectives: linguistic complexity, content foci, and functional move structure. In addition, comparisons were made between edited (by professional writing services) and unedited CEAL-PSs to examine whether the editing process significantly changed the texts of CEAL-PSs in the above three perspectives.

2.4 Corpus Linguistics and Genre Analysis

In contrast to the reviewed studies, the current study adopted corpus linguistics methods as the primary research design in order to enhance the reliability of data analysis. The term "corpus" refers to "an electronically stored collection of samples of naturally occurring language" (Huston, 2006, p. 234). For instance, the American National Corpus (ANC) contains 22 million words of written and spoken data collected since 1990s (Reppen, Ide & Suderman, 2005). Another widely used corpus is the British National Corpus (BNC), which is a collection of 100 million words of written and spoken language samples from a wide range of sources collected in the late 20th century (Huang, 2011). Using corpora to investigate language patterns is referred as "corpus linguistics."

The goal of corpus linguistics is to investigate language patterns by analysing a large quantity of language samples collected from natural settings (Biber, Conrad, & Reppen, 1998). Corpus linguistics studies have developed very quickly in the last three decades. It has been widely applied to explore language variations, metaphor usage, and discourse structures. A comprehensive review of contemporary corpus linguistics applications can be found in Baker (2009).

In general, there are two different approaches of corpus linguistics research: corpus-driven and corpus-based studies. The corpus-driven approach focuses on the discovery of language patterns through exploratory analyses of self-composed corpora whereas the corpus-based approach focuses on the test of researchers' hypotheses through well-established corpora (e.g. ANC and BNC). The present study adopted the corpus-driven approach since the analyses of CEAL-PSs were essentially exploratory and inductive.

The combination of genre research and corpus linguistics has been proposed since late 1990s (e.g. Flowerdew, 1998, 2005; Henry & Roseberry, 2001; Upton & Connor, 2001). There are three major advantages of using corpus linguistics methods in genre studies. First, it can significantly reduce researcher bias. Subjectivity is a key criticism toward qualitative research methods. As human beings, researchers are impossible to become completely objective since their interpretations of data are inevitably influenced by their existing preferences and preconceptions. Even with a high-level of self-awareness and reliable analysis, researchers' perception is still constricted by subconscious bias and personal circumstances (Baker, 2006). Although using corpora cannot remove researcher bias, it at least minimizes researchers' potential cognitive bias by showing concrete statistical data. For instance, literacy studies traditionally rely on the researchers'

intuition to interpret the themes and foci of literary texts. Now, with the application of corpus methods, the frequent word lists can directly reveal the lexical patterns of texts, which can then enhance the reliability of our interpretations regarding text themes.

Second, corpus linguistics methods are applicable for genre studies due to the incremental effect of discourse (Baker, 2006). As discussed in Section 2.1, a specific genre is a collection of texts sharing common lexico-grammatical and discoursal features. In other words, to discover linguistic features of one particular genre, it is necessary to find repetitive patterns in a large collection of samples belonging to that genre. Corpus linguistics methods, with the application of corpus analysis programs, are particularly good at revealing language patterns by processing large text collections. Unlike traditional qualitative methods that take a top-down process of genre analysis, corpus-driven methods are able to conduct bottom-up analyses, which first detect lexical and grammatical similarities among texts. Thus, depending on the focus of research questions, corpus-driven methods tend to be more efficient than traditional qualitative methods, especially when dealing with large size of text samples. The results generated by corpus linguistics methods are at least as valid as those by traditional qualitative methods (Henry & Roseberry, 2001; Upton & Connor, 2001).

Third, corpus linguistics methods provide data triangulation, which improves the validity of genre studies. “Triangulation” in social sciences refers to the application of multiple analysis approaches or the cross verification of various forms of data during data analysis processes (Newby, 1977). It has been widely accepted by scholars that triangulation is a preferred method to improve the validity of data analysis. There are several advantages of using triangulation during data analysis. In general, triangulation:

(1) is able to combine both qualitative and quantitative methods during data analysis, (2) provides robust interpretations of research questions by offering various empirical data, and (3) anchors research questions from multiple perspectives (Baker, 2006). Most genre studies are qualitative and only explore discourse patterns based on macro analyses of a small amount of texts belonging to one genre (Biber, Csomay, Jones, & Keck, 2004). The triangulation offered by corpus linguistics methods can be beneficial for genre studies in that they can offer alternative perspectives for data analysis, showing results omitted by qualitative methods.

Meanwhile, the applicability of corpus-driven methods for genre analysis has received some strong worded critiques. One major concern of these critiques is the applicability of corpus in genre studies. For instance, Swales (2002) argued that interpreting genre structures essentially required a top-down process since the major purpose of genre studies was to discover macrostructures of texts. Thus, most bottom-up corpus analysis methods, such as lexical/phrasal frequency and concordance, are incompatible with genre studies. Another criticism of corpus-driven genre analysis is the de-contextualization of many large corpora. Contextual information of texts is crucial for genre studies since it assists the interpretation of discourse communities and communicative purposes attached to texts. Due to technical constrains and applicability considerations, most large corpora, (e.g. ANC and BNC) exclude contextual information of the collected text samples. Widdowson (1998, 2002) regarded corpus data as samples of language rather than authentic language since it was extracted from original texts with contextual information deleted. Hunston (2002) shared the same concern and she proposed the absence of social

contexts in corpus analysis (e.g. lexical/phrasal frequency) as the most severe drawbacks of applying corpus linguistics methods in genre research.

All the above critiques are valid and offer valuable considerations when conducting corpus-driven genre studies. However, these drawbacks of corpus linguistics methods can be improved with some modifications in research designs. In general, there are two major approaches to improve the validity of corpus-driven genre studies (Flowerdew, 2005). The first is using self-compiled, specialized corpora with linguistic tags at the discoursal level. Newly developed content analysis programs (e.g. NVivo) allow tagging functional moves in corpora, which realizes the traditional hand tagging process in a corpus environment. Thus, if the corpora used for genre studies allow for the tagging practices, the de-contextualization problem of corpus analysis can be solved by the researcher's active involvement during coding processes. Several previous studies have adopted such strategy during their corpus preparation procedures and their results have suggested the effectiveness of introducing discourse tagging systems to corpora (Henry & Roseberry, 2001; Upton & Connor, 2001; Hyland & Tse, 2009).

In terms of the conflict between bottom-up and top-down analysis processes, Flowerdew (2005) suggested using follow-up qualitative methods (e.g. interview, survey, qualitative text analysis) to validate the findings generated by corpus-driven methods. Such strategy has also been used by several previous studies. For instance, Hyland (1998) consulted 80 specialist informants to verify his findings on hedging device usage in research articles. Similar procedures were also conducted in Brown (2004) and Samraj and Monk (2008).

To conclude, with rigorous research designs, corpus linguistics methods can be an effective tool for genre research in discovering language patterns of less studied genres. In the current study, a corpus-driven methodology design was adopted to investigate the collected CEAL-PSs.

Chapter 3 Methodology

3.1 Research Questions

The present study investigated three research questions, which were formulated based on the gaps in previous studies of the PS genre.

1. What is the linguistic complexity of PSs written by Chinese EAL applicants? Is there any inter-disciplinary variation? Do PSs edited by professional writers differ from the unedited ones in their linguistic complexity levels?
2. What are the content foci in PSs written by Chinese EAL applicants? Is there any inter-disciplinary variation? Do PSs edited by professional writers differ from the unedited ones in their content foci?
3. What is the general functional move structure of PSs written by Chinese EAL applicants? Is there any inter-disciplinary variation? Do PSs edited by professional writers differ from the unedited ones in their functional move structure?

3.2 Data Collection

Two types of PS samples were used in the present study: unedited PSs by Chinese EAL applicants (CEAL-PSs) and edited ones (E-PSs). A total of 214 PS samples were collected, including 182 CEAL-PS and 32 E-PS samples. Then, a total of 100 CEAL-PS and 20 E-PS samples were further selected for the data analysis. Six corpora were compiled for the data analysis, including five CEAL-PS corpora and one E-PS corpus. Each CEAL-PS corpus included 20 CEAL-PSs from each of the following five disciplines (Arts, Business, Social Sciences, Engineering, and Sciences). By contrast, the corpus for E-PSs included 20 PSs of various disciplines, with two to five samples in each discipline.

3.2.1 The Collection CEAL-PS Samples

The CEAL-PS samples for the present study were collected from the online forum Gter (Gter, n. d.). Gter is one of the largest online forums in China with an exclusive focus on topics related to Western university applications. It has more than 1 million registered users and approximately 1,700 new posts every day. In China, many prospective applicants for Western graduate schools browse Gter to discuss issues regarding graduate school application, such as advice for program selection, PS writing, and language test preparation. The large visitor volume of Gter makes it an ideal resource to recruit participants for the current study.

To collect the CEAL-PS samples, an invitation letter (see Appendix A) was posted in five major sections of Gter ("American Graduate School Applications," "Canadian Graduate School Applications," "Hong Kong Graduate School Applications," "British Graduate School Applications," and "Australian/New Zealand Graduate School Applications") to invite current Western graduate school applicants from China to submit their PSs for the present study. An implied consent form was attached to the invitation letter. Applicants for graduate schools in Hong Kong were also included in the CEAL-PS collection process as PSs written in English were also required by universities in Hong Kong. In addition, to ensure that the selected CEAL-PS samples were all written by advanced EAL learners, participants who sent in their PSs were asked two questions regarding their English proficiency:

1. Have you passed the minimum English proficiency requirements set by the graduate programs which you are applying for?
2. Could you please provide your latest total TOEFL or IELTS score (you can provide either the exact score or a general range such as 80+ and 100+)?

The final compiled CEAL-PS corpora for data analysis only included samples by participants who answered "Yes" to both questions. In other words, the final selected CEAL-PS samples were all produced by writers who had achieved at least 6.5 in IELTS or 80 in TOEFL. A total of 182 CEAL-PS samples were collected, and 100 of them were selected for the formal analysis, with 20 each in five major academic disciplines (Arts, Business, Social Sciences, Engineering, and Sciences). The length of these CEAL-PS samples ranged from 600-1,200 words due to different graduate programs' requirements. PS samples that were shorter than 600 words or longer than 1,200 words were excluded from the formal analysis.

Table 6 shows the academic disciplines of the 100 selected CEAL-PS samples. It should be noted that the five academic disciplines of CEAL-PSs are further grouped into two divisions: Arts and Sciences. Such divisions come from the academic classification of the Chinese secondary education system (Dello-Iacovo, 2009), in which non-scientific academic courses are generally regarded as "Wen-Ke" (Arts) and scientific courses are generally regarded as "Li-Ke" (Sciences). This system extends to the tertiary level in China and is therefore used in the present study.

Table 6 ***Disciplinary Information of Selected CEAL-PS Samples***

Discipline		Major and sample numbers
The Arts Division	Arts	Applied Linguistics (5), Arts Administration (1), Asian Studies (8), Comparative Literature (1), Cultural Studies (2), Translation (3).
	Business	Accounting (3), Advertising (2), Corporate Communication (2), Finance (3), Hotel Management (2), Management (2), Public Administration (3), Real Estate (3)
	Social Sciences	Anthropology (2), Communication (3), Economics (2), Environmental Policy (1), Journalism (2), New Media (1), Political Sciences (2), Psychology (2), Public Health (1), Public Relation (1), Social Work (2), Sociology (1)
The Sciences Division	Engineering	Chemical Engineering (3), Civil Engineering (3), Computer Engineering (3), Electronic Engineering (3), Information Technology (1), Logistics (3), Manufacture Engineering (2), Mechanic Engineering (2)
	Sciences	Biology (4), Chemistry (3), Computer Science (4), Environmental Studies (4), Material Science (2), Physics (2), Plant Science (1)

Note. The numbers in the bracket refer to the number of PS samples in one discipline. In China, Applied Linguistics is considered as a major in Humanities.

All the selected CEAL-PS samples were transferred into the plain text format (.txt) for corpus compilation. Their original paragraph structures and typos were kept. With the application of the corpus tool Corpus Builder[2], five corpora were compiled and each of them contained 20 PS samples in one Discipline: the Arts Corpus, the Business Corpus, the Engineering Corpus, the Sciences Corpus, and the Social Sciences Corpus.

3.2.2 The Collection of E-PS Samples

In the present study, the term "edited PSs" (E-PSs) refers to PSs written by graduate school applicants from China, and then polished by professional editors. These E-PSs are often posted online as advertisements for attracting potential customers to use the editing services. All essay editing services claim that their posted PSs have helped their clients

[2] Corpus Builder: http://www.lextutor.ca/tools/corpus_builder2/

successfully enter their desired programs. Although such claims are difficult to verify, these E-PS samples have been used in Ding (2007) as successful PS samples. Data analyses of previous studies have shown that E-PSs in general are carefully worded, well organized, and thus can be regarded as successful PS samples (Brown, 2004; Ding, 2007).

In line with previous studies, E-PS samples for the present study were collected from websites of online essay editing services (see Appendix B for details). All of the collected E-PS samples were first written by Chinese EAL applicants, as indicated by the demographic information either on the posting websites or in the texts. A total of 32 E-PS samples were collected and then 20 of them were selected to compile the E-PS Corpus (for the selection criteria, see Section 3.2.1).

As shown in Table 7, E-PS samples in the E-PS Corpus came from five disciplines. Table 8 provides the token numbers of the compiled CEAL-PS and E-PS corpora.

Table 7 ***Disciplinary Information of Selected E-PS Samples***

Discipline	Major and number of PS samples
Arts	Language Education (1), Literature and Culture (1)
Business	Accounting (2), Finance (1), MBA (2)
Engineering	Civil Engineering (1), Computer Engineering (1), Electronic Engineering (1), Mechanic Engineering (2)
Sciences	Biology (1), Chemistry (1), Computer Science (1), Environmental Studies (1), Medical Science (1)
Social Sciences	Communication (1), Economy (1), Public Health (1)

Note. The numbers in the bracket refer to the number of PS samples in one discipline.

Table 8 ***Token Size of the Complied Corpora***

Corpus Name	Sample No.	Token Size
Arts Corpus	20	19532
Business Corpus	20	18917
Engineering Corpus	20	18507
Sciences Corpus	20	18552
Social Sciences Corpus	20	20002
E-PS Corpus	20	17330

Note. "Token size" refers to the total number of words in a particular corpus.

3.3 Data Analysis

Six analyses were conducted on the compiled corpora in the present study (Table 9). These analyses were conducted from three perspectives: (1) linguistic complexity (i.e. the lexical diversity and readability analyses), (2) content foci (the lexical frequency, collocation, and concordance analyses), and (3) textual structure (the functional move analysis). In addition, potential distinctions between unedited and edited PSs were also investigated from each perspective.

The present study adopted three corpus linguistics programs for the data analysis: Range (Heatley, Nation, & Coxhead, 2002), Readability Analyser (Xu & Jia, 2009), WordSmith Tools (Scott, 2008). Two data analysis programs were used for analysing the qualitative and quantitative data generated during the data analysis: NVivo 8 (QSR International Pty, 2008), and SPSS 19 (IBM Corp., 2010).

Table 9 ***Corpus Analyses Conducted on the Compiled Corpora***

Research Question	Analysis	Analysis Software
(1) Linguistic Complexity	• Lexical Diversity • Readability	Range Readability Analyser
(2) Content Focus	• Lexical Frequency • Collocation Frequency • Key-Word-In-Context (KWIC)	WordSmith Tools
(3) Functional Move	• Functional Move Structure	NVivo

The general procedure of each analysis in Table 9 was as follows. One analysis (e.g. the lexical frequency analysis) was first conducted on each compiled corpus. Then the results were compared with each other in search of inter-corpus differences. The comparisons across the CEAL-PS corpora indicated inter-disciplinary variations, whereas the comparison between CEAL-PS corpora and the E-PS corpus explored whether the editing process significantly changed the texts of unedited CEAL-PSs.

The following sections describe the six analyses and their operation procedures.

3.3.1 Linguistic Complexity Analyses

The lexical diversity and the readability analyses were conducted on the compiled corpora to investigate their linguistic complexity.

3.3.1(a) Lexical diversity. The lexical diversity analysis compares the word lists of target corpora (corpora being analysed) with reference word lists (typical words of a particular genre) and shows how many lexicons in the reference word lists are covered by the target corpora. Results of this analysis indicate the target corpora's degree of lexical richness of the target corpora and their authors' language proficiency levels (Heatley, Nation, & Coxhead, 2002).

The present study used Range (Heatley, Nation, & Coxhead, 2002) for the lexical diversity analysis. Range is a lexical analysis program used to detect the lexical distribution of target corpora. It analyses the lexical distribution of target corpora at different lexical frequency levels by comparing the word lists of the target corpora with its reference word list(s). The validity of Range has been verified by its applications in several previous corpus-driven lexical studies, such as Nation (2004) and Wang and Nation (2004).

In the present study, the compiled corpora were imported into the program Range and their word lists were generated by the program and then compared with its default reference word list. The results of each corpus were presented as lexical distribution figures in four categories (Level One, Level Two, Level Three, and Other, explained in the following paragraphs, which indicated how much of the imported corpora was distributed in each category.

Range presents the results of imported corpora in four categories: Level One, Level Two, Level Three, and Other. According to Nation and Coxhead (2001), the four categories are arranged according to lexical frequencies: Level One includes the most frequent 1000 words (the base and derived forms) of English; Level Two includes the second 1000 most frequent words; and Level Three includes words not in the first 2000 words of English but which are frequent in upper secondary school and university texts from a wide range of subjects; the "Other" category refers to lexicons that are not included in the previous three levels.

The default reference word list of Range is produced from the General Service List (GSL) (West, 1953) and the Academic Word List (AWL) (Coxhead, 2000). GSL includes 2000 lexicons, which represents the most frequently used English lexicons in general English. AWL includes 570 word families that frequently appear in academic English writing. AWL does not include the lexicons listed in GSL.

3.3.1(b) Readability. The readability test estimates the overall reading difficulty of texts in target corpora by calculating their Flesch Reading Ease scores (FRE) (Flesch, 1948) and Flesch–Kincaid Grade Level scores (FKGL) (Kincaid, Fishburne, Rogers, & Chissom, 1975). FRE and FKGL tests use the same criteria of evaluation (total words, total sentences, and total syllables), but with different weighting factors. The FRE scores range from 0-100, and a low FRE score indicates difficult reading materials. By comparison, the FKGL scores range from 1-22, which translate the original FRE scores into the U. S. grade levels. For instance, U.S. high school graduates are expected to be able to comprehend texts in grade 17-18 and the standard for people completed tertiary

education is grade 18-22 (Kincaid et al., 1975). Thus, a low FKGL score indicates easy reading materials. The formulas of these two tests are shown in Figures 1 and 2.

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

Figure 1. **The Flesch reading ease test (Flesch, 1948)**

$$0.39\left(\frac{\text{total words}}{\text{total sentences}}\right) + 11.8\left(\frac{\text{total syllables}}{\text{total words}}\right) - 15.59$$

Figure 2. **The Flesch–Kincaid grade level test (Kincaid et al., 1975)**

The present study adopted Readability Analyser (Xu & Jia, 2009) for the readability analysis. Readability Analyser is a text analysis program released by Chinese National Research Centre for Foreign Language Education. This program is able to conduct several typical readability tests on text samples to measure their readability levels, including the FRE and FKGL tests.

In the present study, FRE and FKGL scores of PS samples in each compiled corpus were calculated by Readability Analyser. Then, the scores of each compiled corpus were imported into SPSS 19 and compared with each other via the Kruskal-Wallis test. The Kruskal-Wallis test is a non-parametric test used in this study to compare independent variables.

3.3.2 Content Focus Analyses

Three analyses were conducted to investigate the content foci of the compiled corpora at the lexical, phrasal, and discoursal levels: the lexical frequency analysis, the collocation analysis, and the concordance analysis.

The present study adopted WordSmith Tools (Scott, 2008) for the content focus analyses. WordSmith Tools is a software package for corpus linguistics analysis. It is a

collection of several corpus analysis tools (e.g. Lexical Frequency Calculator, Collocation Analyser, and Concordance Analyser). This software is able to conduct most of typical corpus analyses on target corpora to search for language patterns, such as lexical frequency, collocation, concordance, and so on. WordSmith Tools is one of the most popular corpus linguistics software packages in contemporary corpus linguistics research, and its high validity has been supported by many previous studies (e.g. Berber Sardinha, 2005; Scott, 2001).

3.4.2(a) Lexical frequency. The term "lexical frequency" refers to word frequency counts and percentages that each word makes toward the corpora, which is generally presented as word frequency lists. Lexical frequency lists often provide information of the general lexical preferences and patterns of target corpora.

In the present study, word frequency lists of each compiled corpus were generated by WordSmith Tools, and these lists were compared with each other to identify potential inter-corpus variations at the lexical level.

3.4.2(b) Collocation. The phenomenon that certain words often co-occur with each other is defined as "collocation" (Baker, 2006, p. 96). The collocations of a target word can provide valuable information of the phrasal patterns in target corpora.

In the present study, the collocation analysis was conducted on key lexicons identified in the lexical frequency analysis by WordSmith Tools. These key lexicons' collocations lists in each compiled corpus were compared with each other to identify potential inter-corpus variations at the phrasal level.

3.4.2(c) Concordance. According to Baker (2006), "concordance" refers to "a list of all of the occurrences of a particular search term in a corpus, presented with the contexts

that they occur" (p. 71). The concordance analysis is one of the most frequent analyses for corpus linguistics research. Figure 3 shows how the KWIC concordance is conducted using the WordSmith Tools. When a key word is targeted, all the sentences containing the word are located and listed with the key term in the central position.

Hit	KWIC	File
1	eone who is focused on furthering my education, I would like to a	WPS-Account:
2	na, a ¡°211¡± project university. My goal is to further my studie	WPS-Account:
3	university. My goal is to further my studies in more economically	WPS-Account:
4	heory so that I can contribute to my motherland in the future. It	WPS-Account:
5	l Risk Analysis¡±are suitable for my pursuits. I am excited about	WPS-Account:
6	ial Risk Mmanagement. During my Bachelor studies, I emphasiz	WPS-Account:
7	ture learning. In several of my classes, I served as the lea	WPS-Account:
8	Professor Hu, was impressed with my unique presentation. I	WPS-Account:
9	in many social activities during my studies. During the summer o	WPS-Account:
10	hers for this competition. During my stay in ChangSha, I connecte	WPS-Account:
11	any manager was so satisfied with my work on the report that he o	WPS-Account:
12	ned to the Red Cross and to share my experiences with the media.	WPS-Account:
13	ences with the media. During my school life, I was elected c	WPS-Account:
14	t or auditor in China and to have my own accounting firm is my ma	WPS-Account:
15	to have my own accounting firm is my main goal within the next te	WPS-Account:
16	nting and financing department of my father¡¯s XYZ Corporation. T	WPS-Account:

Figure 3. **The screenshot of a sample concordance analysis**

In the present study, key lexicons identified in the lexical frequency analysis were investigated across the compiled corpora by WordSmith Tools. The concordance lines of the key lexicons were located and then grouped into different categories according to their common themes. The results of each compiled corpus were compared with each other to identify potential inter-corpus variations at the discoursal level.

3.4.3 Functional Move Analysis

The functional move analysis was conducted on the compiled corpora to investigate their textual structures. The present study adopted NVivo for coding the functional moves of the selected PS samples. NVivo 8 (QSR International Pty Ltd, 2008) is a Qualitative Data Analysis (QDA) software package, which allows users to categorize and analyse materials (audios, videos, pictures, and texts) with various codes. It also includes general

text processing functions, such as extracting specific text, searching target word/phrases, and adding comments to original texts.

In the present study, functional moves of each selected PS sample were coded at the clause level. For instance, if there were two clauses in one complex sentence discussing two different topics, they were coded as two separate moves respectively. On the other hand, if there were three paragraphs in one PS sample discussing the same topic, they would be coded only as one move.

The coding scheme was adapted from previous studies with minor modifications (Ding, 2007; Samraj & Monk, 2008). I combined the coding schemes in the previous studies (Tables 4 and 5) based on their commonalities and then tested the combined coding scheme by coding 20 PS samples in the collected data. The finalized coding scheme used for the current analysis is shown in Table 10.

Table 10 ***Functional Moves of PS Coded in the Current Study***

General Moves	Sub Moves
Move 1: Introduction	
	Generalization/Attention Catcher
	Background
	Reasons for Application
	Goals or Decision to Apply
Move 2: Establishing credential	
	Education
	Research
	Work
	Personal Attributes
	Other Experience
Move 3: Reasons for application	
	Gap in Background
	Positive Gains (incl. Interests)
	Program/University Attributes
	Disciplinary and Research Reasons
Move 4: Conclusion	
	Goals/Predictions of Future
	Self-evaluations
	Final Appeal

Adapted from Ding (2007) and Samraj and Monk (2008)

To limit the influence of researcher's bias, secondary verification methods are required by qualitative analysis methods (Seliger & Shohamy, 1989). To improve the reliability of the move coding procedure, intra-coder reliability was conducted by the current analysis. The moves of the collected PS samples were coded twice, with a two-week interval. The results of the two coding practices were compared and an agreement rate of 97.55% was reached between the two coding results. The coding results were further examined after six months to ensure the consistency of the coding results.

The definitions of each general and sub moves in Table 10 are given in the following sections.

3.4.3(a). Move 1: Introduction. This move is primarily used to get readers' attention or provide a brief introduction of the PS content. The introduction section is a common move in academic writing (Swales, 1990). In promotional genres, it is generally realized by the opening paragraphs in promotional genres, such as cover letters written for job applications (Bhatia, 1993). The "Introduction" move includes three sub-moves: "Generalization/Attention Catcher," "Background," and "Goals or Decision to Apply."

- **"Generalization/Attention Catcher"** refers to descriptions that catches reader's attention, or claims the aim of the PS. This sub-move is generally realized as the first sentence of the opening paragraph.
- **"Background"** refers to descriptions of an applicant's education and personal background information. This sub-move is generally realized in the subsequent sentences after the first sentence of the opening paragraph.

- **"Goals or Decision to Apply"** refers to descriptions of explicit reasons for an applicant's graduate school application decision. This sub-move generally closes the introduction section.

3.4.3(b) Move 2: Establishing credentials. This move provides information regarding an applicant's credentials for studying and conducting research at the graduate level. This move has been recognized as an essential move of a PS sample by previous studies (Brown, 2004, Ding, 2007; Samraj & Monk, 2008). The "Establish Credentials" move includes five sub-moves: "Education," "Research," "Work," "Personal Attributes," and "Other Experience."

- **"Education"** refers to descriptions of an applicant's education background and academic achievements. GPA information and completed courses are usually discussed in this sub-move.
- **"Research"** refers to descriptions of an applicant's research experience, research interest, or potential research area of graduate study. Completed research projects and research interests are usually discussed in this sub-move.
- **"Work"** refers to descriptions of an applicant's work/internship experience. This sub-move can be frequently observed in PSs for professional programs (Samraj & Monk, 2008).
- **"Personal Attributes"** refers to descriptions of an applicant's character and personal attributes. This sub-move was regarded as a "peripheral move" in Ding (2007).

- **"Other Experience"** refers to descriptions of an applicant's volunteer activities and other pertinent experience. Clauses not belonging to the four sub-moves above were coded as "Other Experience" in the current analysis.

3.4.3(c) Move 3: Reasons for application. This move discusses an applicant's reasons for applying for a particular graduate program. This move has also been recognized as an essential move by previous studies (Brown, 2004, Ding, 2007; Samraj & Monk, 2008). The "Reasons for Application" move includes four sub-moves: "Background Gap," "Positive Gains," "Program/University Attributes," and "Discipline/Research Reasons."

- **"Background Gap"** refers to descriptions of how an applicant's background can be improved by the desired graduate program. This sub-move is frequent in PSs written for professional programs, according to Samraj and Monk (2008).
- **"Positive Gains"** refers to descriptions of how an applicant's personal interests and career goals can be realized by the desired program. This sub-move usually follows the "background" sub-move to increase the persuasion of an applicant's appeal.
- **"Program/University Attributes"** refers to descriptions of how an applicant is attracted by the desired program/university. This sub-move is the hook section to explicitly connect an applicant with his/her desired program(s).
- **"Disciplinary/Research Reasons"** refers to descriptions of an applicant's desire to pursue a research career after his/her graduate study. This sub-move is very important for research programs (Samraj & Monk, 2008).

3.4.3(d) Move 4: Conclusion. This move summarizes the PS and provides a final appeal for the admission committee's favourable consideration. It is usually realized as the closing paragraph in one PS sample. The "Conclusion" move includes three sub-moves: "Goals/Predictions of Future," "Self-evaluations," and "Final Appeal."

- **"Goals/Predictions of Future"** refers to descriptions of an applicant's future goals upon completion of the desired program.
- **"Self-evaluations"** refers to descriptions of an applicant's final self-evaluation.
- **"Final Appeal"** refers to descriptions of an applicant's final appeal for the admission into the desired program.

Figure 4 shows the coding procedure of one PS sample from the collected data. Following the same procedure, functional moves of each selected PS sample were coded and the frequencies of each move were compared across the compiled corpora. The Kruskal-Wallis test was adopted again to statistically analyse the results. In addition, the average length of some key moves was calculated by NVivo and compared across the compiled corpora.

The Sample CEAL-PS (underlined phrases refer to functional move labels)

Strolling on the street lanes, sitting in the cafeteria or entering an elevator at X University of China is an unforgettable experience. With students from 56 ethnic groups from all over the country, X seems like a miniature community-a veritable kaleidoscope of Asia with students proudly speaking their native languages, yet finding similarities amidst their differences. This has been at times bittersweet for me, for although I grew up bilingual in Mandarin and Cantonese, my parents never spoke to me in our ethnic language of Zhuang. **Background** In fact, my fascination with Second Language Acquisition actually originated from this sad fact of my language loss. I grew up straddling two worlds, modern Mandarin dominated China, and the ever-dwindling Zhuang traditions. I have watched the Zhuang culture eroding with each passing year, and I long to research how modern and traditional languages could co-exist. **Reason for application**

My interests, therefore, have settled on teaching, as I see this path as the most effective in promoting cultural tolerance and cross-cultural and linguistic understanding. In my experiences teaching Chinese English learners, and American and Thai Chinese learners, I have encountered problems in motivation, attitudes, pedagogical materials, and teaching methods. I aim to become a teacher and researcher who is able to understand and solve the aforementioned problems. I wish

to receive systematic training from MA-PhD Linguistics Program with concentrations on Applied Linguistics, especially on Second Language Acquisition. My preparation for this program includes three parts: Academic Studies, Research experiences, and Teaching experiences. **Positive gains**

In the past five years, I have gained extensive knowledge and training for both English and Chinese linguistics, Second Language Acquisition as well as intercultural studies which give me a critical insight into English learning from both linguistic and cultural aspects. To further nourish my academic interests, I have also audited lectures like Global English, English Learning strategies, Corpus linguistics, and Textual Analysis, and become familiar with the latest topics and methodologies in language learning. However, while undertaking my project and writing my undergraduate thesis, I realized that the knowledge of L1 is also vital for L2 instruction and research. To better my understanding in this area, I decided to apply for the Volunteer Chinese Teaching Program as I graduated and has received comprehensive trainings on Chinese linguistics and Culture and L2 Teaching pedagogies. All the knowledge I have acquired will give me an edge in future studies of Chinese speakers learning English as a second language, or vice versa. **Education**

In addition to classroom studies, I also enhanced my research abilities and put what I have learnt into practical application. To describe, in the Undergraduate Research and Training Program I undertook to solve Chinese college students' problems in constructing fluent sentences in English. The team I organized worked on a research project on C-E interpreting training based on the frequently-used sentence patterns. We have compiled a self-study textbook to aid formal interpreting training, which begins with a brief introduction of frequently-used sentence patterns, ensuing are analysis of translation strategies, training schedule, translation exercises and vocabulary instruction. My consistent interest in learner's performance also led to my graduation thesis: Translation Strategies of Cohesive Devices in C-E Interpretation of English Majors: A Case Study on Repetition. All these experiences have sharpened my research ability and consolidated my career plan as a teacher and researcher. **Research**

Apart from academic achievement, I have also explored every opportunity to study second language teaching and learning. My most valuable experience has been teaching Chinese to Thai students in Thailand on a volunteer basis, while simultaneously learning Thai. My classes at a local Thai government school have consisted of more than forty learners each period. I have found through trial and error that studying through games and tasks-based language teaching techniques would be the most useful way to enhance students' interest and efficiency in learning. Discussing with and learning from my supervisor Mrs. X, I have devised games like "Sequent Bubbles" to drill students' pronunciation, adapted songs to teach sentence patterns, and turned my classroom into a "supermarket" when teaching fruits. It turned out that as high as 92 percent of students could match the word with correct meaning, as compared with the traditional teaching methods. **Work**

I feel incredibly lucky to have the opportunity to regain Zhuang language and gain fresh ideas on second language acquisition through my own studies of the Thai language. The power of context in L2 immersion leads me to topics like the cognitive process of producing a second language, for which I look forward to studying at the University of X. I want to research with what amount of input learners could solve problems in second language scenarios, and what kind of pedagogical materials can be constructed using our findings. I have observed that bilingualism teems in every aspect of Thai life through ASEAN's policy to promote English learning. It has prompted my interest in conducting a comparative analysis on English education in Asia and the role of target language teacher (foreign language teacher) in integrating culture with language learning.

Disciplinary & research reasons

Last but not least, with my multi-lingual experience, I feel interested to investigate the differences in Chinese-English and English-Chinese language acquisition. My short-term career goal is to become a well-trained teacher in English and/or Chinese to Speakers of Other Languages. And my long-term aspiration is to launch a well-organized program in helping learners regain their desired identity. **Positive gains** To pursue this goal, I first need to gain a solid theoretical foundation in both linguistic theory and teaching practices. The MA-PhD program in Linguistics at University of X emphasizes both theory and application and would be a perfect research environment for me to fulfill my goals and explore the social impact on learning a foreign language. **Program/university attributes**

I prefer empirical studies by using large corpora and basing my analysis on substantial data. I have learned some corpus tools like CLAWS, ParaConc and WordSmith by self-study and I would like to be supervised by and work with Professor X, Director of X Language Media Center, to lay a solid foundation in computational linguistics and corpus linguistics, especially for educational instruction. Under her instruction, I would be able to develop effective pedagogical materials by deducing frequently-used phrases and syntactic structures from authentic texts, identify learners' strategies and detecting learners' errors in learning. **Disciplinary/research reasons**

At last, I look forward to collaborating and researching alongside the University of Pittsburgh's distinguished faculty and students towards better understanding Linguistic theory and application. **Final appeal**

Figure 4. **One coded CEAL-PS sample**

Chapter 4 Results

The data analysis results are presented in three sections, organized according to the research questions. Section 4.1 provides information on the lexical diversity and the readability analyses. Then, the lexical frequency, the collocation, and the concordance analyses are presented in Section 4.2. Finally, Section 4.3 shows the results of the functional move analysis.

4.1 Research Question 1: Linguistic Complexity of the Collected CEAL-PSs

The linguistic complexity of CEAL-PSs was investigated from two perspectives: lexical diversity and readability. In general, the two analyses showed similar linguistic complexity results across the compiled corpora[3], with some slight distinctions between corpora in the Arts and the Sciences Divisions.

4.1.1 Analysis of Lexical Diversity

The lexical diversity analysis results of the compiled corpora are presented in Tables 11 and 12. Table 11 shows the results as token quantities whereas Table 12 shows them as lemma quantities. Both tables present the compiled corpora's lexical distributions in four categories: Level One, Level Two, Level Three, and the "Other" category. The three levels are arranged according to the lexical frequencies of Range's default reference word list. Level One includes words with the highest frequencies in the default word list, whereas Level Three includes words with the lowest frequencies. The category "Other" is used for words that are not in the default reference word list.

[3] The compiled corpora refer to: the Arts Corpus, the Business Corpus, the Social Sciences Corpus, the Engineering Corpus, the Sciences Corpus, and the Edited-PS Corpus.

In general, the lexical diversity results of each corpus across the four categories were identical. As suggested by Table 11, each corpus included approximately 73% tokens at Level One, 5% at Level Two, 10% at Level Three and 10% in the Other category. According to Table 12, each corpus had approximately 40% lemmas at Level One, 12% at Level Two, 17% at Level Three, and 31% in the Other category. Thus, the majority of lexicons used by the compiled corpora fell into Level One, which includes the most frequent lexicons in academic writing (Heatley, Nation & Coxhead, 2002).

Nonetheless, several slight distinctions emerged between the Arts Division (the Arts Corpus, the Business Corpus, and the Social Sciences Corpus) and the Sciences Division (the Engineering Corpus and the Sciences Corpus). As shown in Tables 11 and 12, corpora in the Arts Division tended to have higher percentages of lexicons at Level One than corpora in the Sciences Division. In Table 11, the Arts Corpus had the largest percentage of lexicons at Level one (76.74%), followed by the Business Corpus (74.69%) and the Social Sciences Corpus (73.48%). By comparison, the Engineering Corpus had fewer tokens (73.17%) at Level One, followed by the Sciences Corpus (72.92%). Table 12 also presents similar rankings. Meanwhile, corpora in the Sciences Division tended to include more lexicons in the "Other" category than corpora in the Arts Division in both tables. In Table 11, figures of the "Other" category for the Engineering and Sciences corpora were 10.85% and 11.20% respectively, in contrast to the figures of the Arts Division (Arts. 9.94%, Business 8.88%, and Social Sciences 10.59%). Such distribution pattern was followed in Table 12. However, it should be concluded that the results showed that the differences between the Arts and the Sciences Divisions in lexical diversity were very small.

The results of the E-PS Corpus in Tables 11 and 12 were similar to those of unedited corpora. In Table 11, the E-PS Corpus had 73.58% at Level One, 5.61% at Level Two, 10.12% at Level Three, and 10.69% in the "Other" category. The results in Table 12 were 39.96%, 10.92%, 17.92%, and 31.20% respectively. Therefore, the data analysis suggested that the edited PS samples were similar to the unedited PS samples in lexical diversity. The *p* values of the Kruskal-Wallis test were 0.416 in Tables 11 and 12, indicating no statistical significance of lexical distributions across the corpora (the *p* values for both tables are identical since both tables present the same lexical distribution results by different measurements).

Table 11 ***Lexical Diversity Analysis Results (Token) of the Compiled Corpora***

	The Arts Division			The Sciences Division		E-PS
	Arts/%	Business%	Social Sciences/%	Engineering/%	Sciences/%	E-PS/%
Level 1	14894/76.74	14032/74.69	14583/73.48	13451/73.17	13450/72.92	12655/73.58
Level 2	968/ 4.99	1073/ 5.71	1056/ 5.32	1153/ 6.27	1130/ 6.13	965/ 5.61
Level 3	1617/ 8.33	2013/10.71	2107/10.62	1786/ 9.71	1799/ 9.75	1740/10.12
Other	1929/ 9.94	1669/ 8.88	2101/10.59	1994/10.85	2065/11.20	1838/10.69
Total	19408	18787	19847	18384	18444	17198

Note. The "Other" category refers to words that are not in the word lists of Range. Kruskal-Wallis test result: p=0.416

Table 12 ***Lexical Diversity Analysis Results (Word Family) of the Compiled Corpora***

	The Arts Division			The Sciences Division		E-PS
	Arts/%	Business/%	Social Sciences/%	Engineering/%	Sciences/%	E-PS/%
Level 1	1348/40.69	1369/40.82	1470/40.08	1394/39.94	1291/38.29	1387/39.96
Level 2	399/12.04	385/11.48	410/11.18	434/12.44	413/12.25	379/10.92
Level 3	546/16.48	637/18.99	631/17.20	584/16.73	539/15.98	622/17.92
Other	1020/30.79	963/28.71	1157/31.54	1078/30.89	1129/33.48	1083/31.20
Total	3313	3354	3668	3490	3372	3471

Note. The "Other" category refers to words that are not in the word lists of Range. Kruskal-Wallis test result: p=0.416

4.1.2 Analysis of Readability

Tables 13 and 14 present the readability analysis results of the compiled corpora. Table 13 shows the compiled corpora's Flesch Reading Ease (FRE) test scores whereas Table 14 shows their Flesch–Kincaid Grade Level (FKGL) test scores. Both tables include each corpus's average, minimum and maximum scores, along with its group range and standard deviation. As explained in Chapter 3, a low FRE score indicates difficult reading material whereas a low FKGL score indicates easy reading material. The Kruskal-Wallis test was applied to the results in both Tables to determine whether there is any statistically significant inter-corpus significance.

In general, the data analysis results suggested similarities across the compiled corpora in readability. In Tables 13 and 14, each corpus achieved similar FRE and FKGL scores. To be specific, in Table 13, the mean FRE scores ranged from 31.32 to 40.28, with the lowest mean score shown in the Social Sciences Corpus (M=31.32) and the highest mean score in the Arts Corpus (M=40.28). The results of the FKGL test had a narrower range from 13.25 to 14.82 in Table 14, with the lowest mean score in the Arts Corpus (M=13.25) and the highest one in the Social Sciences Corpus (M=14.82). The results seemed to indicate that the Social Sciences Corpus was the most difficult to read among the compiled corpora. By comparison, the Arts Corpus and the Business Corpus tended to be easier to comprehend than other corpora. Tables 13 and 14 showed that both corpora had high FRE mean scores (M=40.28 and M=38.57) and low FKGL mean scores (M=13.25 and 13.51).

Meanwhile, there were considerable individual variations among the collected PS samples, as shown by the high standard deviation and range values (Tables 13 and 14).

The standard deviation values for the FRE test range between 9.13 and 10.88 in Table 13 while they range between 1.87 and 2.63 for the FKGL test in Table 14.

The results of the E-PS Corpus were similar to other corpora in Tables 13 and 14. The mean score of the E-PS Corpus was 34.28 for the FRE test and 14.11 for the FKGL test. The similarities across the compiled corpora in both tables were further supported by statistical analysis. The *p* values of the Kruskal-Wallis test were 0.096 in Tables 13 and 0.335 in Table 14, respectively, indicating no statistical significance of readability score distributions across the corpora.

Table 13 ***Flesch Reading Ease Scores of the Compiled Corpora***

	The Arts Division			The Sciences Division		E-PS
Flesch reading ease scores	Arts	Business	Social Sciences	Engineering	Sciences	
Mean	40.28	38.57	31.32	34.43	35.19	34.28
Std. Deviation	9.13	11.13	10.88	9.33	9.73	9.68
Minimum	22.60	11.80	3.90	15.40	15.30	12.50
Maximum	56.50	56.20	49.00	53.40	51.10	54.50
Range	33.90	44.40	45.10	38.00	35.80	42.00
Kruskal-Wallis test result: p=0.096						

Note. Lower Flesch reading ease scores indicate harder reading materials.

Table 14 ***Flesch–Kincaid Grade Level Scores of the Compiled Corpora***

	The Arts Division			The Sciences Division		E-PS
Flesch–Kincaid grade level scores	Arts	Business	Social Sciences	Engineering	Sciences	
Mean	13.25	13.51	14.82	14.35	14.5	14.11
Std. Deviation	1.87	2.25	2.63	2.34	2.37	2.36
Minimum	10.00	9.60	10.90	10.70	10.10	10.00
Maximum	17.40	18.80	21.70	19.80	19.90	20.00
Range	7.40	9.20	10.80	9.10	9.80	10.00
Kruskal-Wallis test result: $p=0.335$						

Note. Lower Flesch–Kincaid grade level scores indicate easier reading materials.

4.2 Research Question 2: Content Focus of the Collected CEAL-PSs

The content foci of CEAL-PSs were investigated at the lexical, phrasal, and discoursal levels. Three analyses were conducted by WordSmith Tools: lexical frequency, collocate, and concordance. The underlying assumption of these analyses is that focused descriptions in the collected PS samples can be indicated by their high frequencies. In general, both similarities and differences were found across the compiled corpora.

4.2.1 Analysis of Lexical Frequency

Tables 15 and 16 present the lexical frequency results of the compiled corpora. Table 15 shows each corpus's top ten lexicons and their frequencies while Table 16 shows each corpus's top ten content words and their frequencies. Content words are defined as nouns, verbs, adjectives, and lexical adverbs in the present study.

As presented in Table 15, most high-ranking lexical items in the compiled corpora were functional words (e.g. determiners, propositions, and pronouns), which did not reveal much valuable information since the frequent usage of functional words is a general feature of English language writing. Nonetheless, results in Table 15 suggested one common feature across the compiled corpora: self-expressions such as "I," "my," and "me" appeared very frequently in Table 15. In particular, the pronoun "I" ranked between top 3 and top 5 in Table 15, even higher than some common prepositions such as "in" and "to."

The results in Table 16 showed both similarities and differences across the compiled corpora. In terms of similarities, general academic lexicons (e.g. "study," "research," and "university") achieved high rankings in the lexical frequency lists across the compiled corpora. Moreover, another commonality across the compiled corpora was the use of

lexicons “China” and “Chinese.” They appeared in most lexical frequency lists in Table 16.

In terms of inter-corpus differences, discipline-specific lexicons only achieved high rankings in their own discipline’s lexical frequency lists in Table 16. For instance, lexicons regarding engineering systems (e.g. “engineering,” “design,” and “system”) only appeared in the Engineering Corpus’s list. Lexicons regarding language and culture (e.g. “English,” “language,” and “culture”) achieved high rankings only in the Arts Corpus’s list. Business-related lexicons (e.g. “management” and “finance”) only appeared in the Business Corpus’s list. Furthermore, the lexicon “research” ranked differently in different corpora. It ranked higher in lists of the Sciences Division, than in those of the Arts Division. For instance, it ranked no.1 in both lists of the Engineering Corpus and the Sciences Corpus, but only no.10 in the list of the Arts Corpus.

The lexical frequency results of the E-PS Corpus in Tables 15 and 16 resembled the results of other unedited PS corpora. The E-PS lists in both tables included general academic lexicons (“research,” “university,” and “study”), the demographic lexicon (“China”), and several discipline specific lexicons (“management” and “project”). Overall, the results of the E-PS Corpus tended to be a combination of general and discipline-specific academic lexicons, which were consistent with the fact that the E-PS Corpus was comprised of PS samples from various disciplines.

The above findings were further confirmed by a follow-up lexical keyness analysis on the compiled corpora (See Appendix D).

Table 15 ***Top Ten Words and their Frequencies in the Compiled Corpora***

	The Arts Division						The Sciences Division				E-PS	
	Arts	Freq.	Business	Freq.	Social Sciences	Freq.	Engineering	Freq.	Sciences	Freq.	E-PS	Freq.
1	the	855	the	892	the	929	the	921	the	972	the	879
2	and	795	and	721	and	744	and	711	and	678	and	658
3	I	647	I	684	in	667	I	657	I	636	I	574
4	to	621	to	659	to	640	of	602	of	631	of	543
5	in	568	in	611	of	638	to	555	to	561	to	535
6	of	557	of	564	I	610	in	552	in	528	in	508
7	my	490	my	427	my	381	my	433	my	449	a	407
8	a	420	a	380	a	352	a	332	a	373	my	402
9	as	230	as	203	with	173	for	187	with	168	for	167
10	for	191	for	193	as	172	that	163	for	161	that	159

Table 16 ***Top Ten Content Words and their Frequencies in the Compiled Corpora***

	The Arts Division						The Sciences Division				E-PS	
	Arts	Freq.	Business	Freq.	Social Sciences	Freq.	Engineering	Freq.	Sciences	Freq.	E-PS	Freq.
1	Chinese	162	social	67	social	93	research	105	research	140	research	87
2	language	113	knowledge	66	China	92	engineering	94	study	80	university	69
3	culture	109	university	64	research	84	university	86	university	75	work	58
4	English	101	study	63	media	70	knowledge	59	engineering	61	school	44
5	teaching	75	management	60	communication	62	study	58	knowledge	61	program	43
6	university	68	China	52	study	61	time	51	project	56	China	41
7	China	66	financial	51	university	60	design	48	program	53	knowledge	39
8	literature	65	research	50	work	59	system	48	science	50	management	36
9	study	65	program	48	Chinese	54	work	47	computer	46	study	36
10	research	64	work	48	school	45	project	46	experience	39	project	35

4.2.2 Analysis of Collocates

As found in the lexical frequency analysis, self-expressions (i.e. "I," "me," and "my") were ubiquitously focused in the CEAL-PSs. Thus, collocates of the pronoun "I" in the compiled corpora were analysed in this section. During the collocation analysis, it was found that many collocates of "I" were also included in the collocations of "me" and "my," since sentences in the collected CEAL-PSs tended to be relatively short. Therefore, collocates of "me" and "my" were not included in the analysis due to the potential repetition problem.

Based on a -5 to +5 span (span refers to searching collocates of a target word within the distance of five lexicons), frequent collocations of "I" were extracted from the collected PS samples. Different forms of one word (e.g. "China/Chinese," "know/knowledge," and "make/made") were calculated together as one lemma. Table 17 shows frequent lemmas collocating with "I" in all collected CEAL-PSs. The table shows top 15 collocates of "I," excluding functional words such as determiners, prepositions, and conjunctions due to the reason described in the previous section.

In general, the results of Table 17 showed that lemmas collocating with "I" can be divided into four categories: (1) mental transitive verbs[4] (i.e. "make," "find," and "want"); (2) academic expressions (i.e. "study," "learn," "work," "research," "knowledge," "believe," "experience," "university," and "school"); (3) chronological expressions (i.e. "time" and "year"); and (4) demographic expressions (i.e. "China"). The appearance of mental transitive verbs in Table 17 was somewhat expected as the major content in PS writing is the applicants' academic background, which inevitably involves gaining

[4] This term derives from Functional Grammar, referring to transitive verbs that describe the interactions between the external world and the internal world (see Thompson, 2004, pp. 92-96 for details).

knowledge from the external world. By contrast, the academic expressions in Table 17, suggested some unique features of the PS genre. Close scrutiny of these academic expressions revealed a narrative framework widely adopted in the collected CEAL-PSs. As shown in Figure 5, this framework described how one's previous experience provides credentials for his/her prospective studies at the graduate level. Finally, the chronological and demographic expressions provided contextual information for the narrative framework. The interconnections between the enlisted lemmas (the dictionary forms of a set of words) in Table 17 were further confirmed by collocation analyses of lexicons "research" and "study" (see Appendix E).

Following the same procedure used in Table 17, frequent lemmas collocating with "I" in each compiled corpus are shown in Table 18. Results of Table 18 presented two major differences between the Arts Division and the Sciences Division. First, the lemma "research" achieved higher rankings in the Sciences Division than the Arts Division in Table 18, showing that research experience and prospects were better described by PSs written for the Sciences Division than those for the Arts Division. Second, the lemma "China" appeared in the lists of the Arts Corpus and the Social Sciences Corpus, but not in those of the Sciences Division, indicating that China-related issues were often mentioned by applicants of the Arts Division.

Finally, results the E-PS Corpus in Table 18 showed a combination of lemmas from the Arts Division and the Sciences Division. On the one hand, many of its enlisted lemmas can be found in the lists of the Sciences Division with similar rankings (e.g. "research" and "project"); on the other hand, the E-PS Corpus included the lemma "China" as well, which only appeared in the lists of the Arts Corpus and the Social Sciences Corpus.

Table 17 ***Frequent Lemmas Collocating with "I" of All Collected PS Samples***

Ranking	Lemma	Freq. (Total)	Freq. (Left)	Freq. (Right)
1	Study	274	164	110
2	Learn	258	79	179
3	Work	214	90	124
4	Research	203	128	75
5	Year	193	162	31
6	Knowledge	193	90	103
7	Believe	173	42	126
8	University	153	114	39
9	China	139	100	39
10	Time	135	107	28
11	Experience	127	80	47
12	Make	127	33	94
13	Find	119	14	105
14	Want	116	9	107
15	School	109	75	33

Note. Different forms of one lemma (e.g. learn and learned) are counted together in this list.

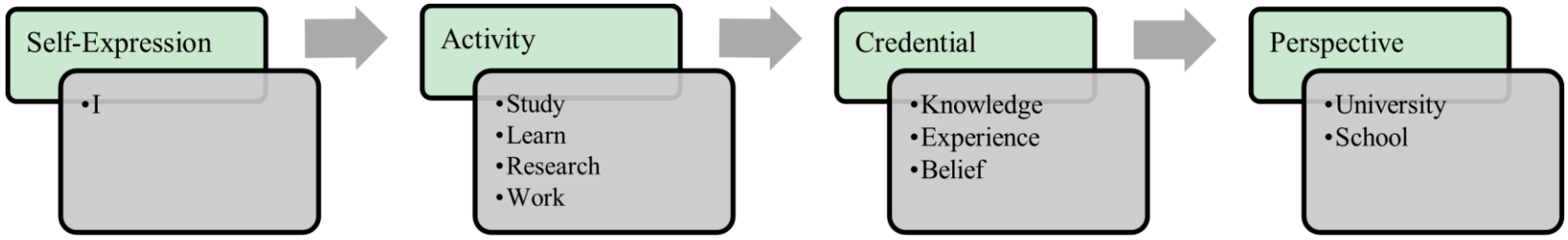

***Figure 5*. The narrative framework of the PS genre**

Table 18 ***Frequent Lemmas Collocating with "I" across the Compiled Corpora***

	The Arts Division						The Sciences Division				E-PS	
	Arts	Freq.	Business	Freq.	Social Sciences	Freq.	Engineering	Freq.	Sciences	Freq.	E-PS	Freq.
1	Study	75	Study	61	Study	62	Study	47	Study	68	Study	62
2	China	57	Learn	55	Believe	33	Research	45	Research	41	Research	36
3	Learn	53	Work	55	Work	33	Learn	41	Learn	36	Work	35
4	Teach	47	Knowledge	51	Learn	31	Work	35	Make	35	Year	32
5	Knowledge	34	Believe	39	Year	29	Engineer	33	Year	35	Knowledge	28
6	English	32	Year	39	Research	28	University	33	Knowledge	33	Learn	27
7	Research	31	Make	34	China	26	Year	33	University	30	China	23
8	Believe	30	Program	30	School	25	Find	28	Find	26	Believe	22
9	Experience	27	Time	28	Knowledge	23	Time	27	Project	26	Interest	20
10	Culture	23	Want	26	Time	23	Project	26	Get	24	Project	20

4.2.3 Analysis of Concordance Lines

Concordance lines of "study," "research," "Chinese," and "China" in the compiled corpora were extracted and analysed by WordSmith Tools. These lexicons were selected due to their unique status in the collected CEAL-PSs. The lexicons "study" and "research" represented two significant aspects of academic activities in the narrative framework of PS writing. "Chinese" and "China" were targeted since they reflected the demographic origin of the PS authors in the present study.

Concordance lines of each lexicon were categorized according to their common themes. Table 19 shows the five theme categories and sample concordance lines. Among those, three categories were used for coding concordance lines of "study" and "research": "Credential," "Prospect," and "Other." Concordance lines of "China" and "Chinese" were also coded with three categories: "Background," "Culture and Society," and "Other." The category "Credential" includes concordance lines that describe applicants' credentials in either study or research aspects. By comparison, the category "Prospect" includes concordance lines that describe applicants' prospects of graduate study in academic and professional developments. Similarly, the category "Background" refers to concordance lines describing the applicants' relevant Chinese background (e.g. hometown, education, and so on). By contrast, the category "Culture and Society" refers to concordance lines discussing social and cultural issues in China. Finally, the category "Other" is used for concordance lines not belonging to the above four categories.

Table 19 ***Concordance Categories and Sample Lines***

Concordance Category	Example
Credential	• My study in undergraduate school was directed and efficient, though suffering sometimes.
Prospect	• I can perform simulation and research once I join any research projects.
Background	• Before I entered RUC, I lived in Nanning, a small city in China.
Culture and Society	• Don't we all wish the Chinese society an opener and charmer place to realize common good?
Other	• In my opinion, the Chinese University of Hong Kong is a bilingual and multicultural institution of higher learning.

Tables 20 to 22 show the concordance analysis results in each compiled corpus. These results suggested several similarities across the compiled corpora. First, the analysed corpora had similar distributions of concordance lines in Tables 20 and 21. Most corpora showed an even distribution of concordance lines between the categories "Credential" and "Prospect" in Table 20 while in Table 21 they contained substantially more concordance lines in the "Credential" category than the "Prospect" category.

Table 20 ***Concordance Lines of "Study" across the Compiled Corpora***

		The Arts Division			The Sciences Division		E-PS
	Total	Arts	Business	Social Sciences	Engineering	Sciences	
Credential	175	31	31	27	31	37	18
Prospect	174	33	28	30	24	41	18
Other	14	1	4	4	3	2	0
Total	363	65	63	61	58	80	36

Table 21 ***Concordance Lines of "Research" across the Compiled Corpora***

		The Arts Division			The Sciences Division		E-PS
	Total	Arts	Business	Social Sciences	Engineering	Sciences	
Credential	338	40	38	50	66	83	61
Prospect	163	23	8	32	35	42	23
Other	29	1	4	2	4	15	3
Total	530	64	50	84	105	140	87

Table 22 ***Concordance Lines of "Chinese/China" across the Compiled Corpora***

		The Arts Division			The Sciences Division		E-PS
	Total	Arts	Business	Social Sciences	Engineering	Sciences	
Background	250	130	27	29	12	25	27
Culture & Society	266	77	32	96	18	16	25
Other	88	21	15	21	7	15	9
Total	604	228	76	146	37	56	61

Moreover, content analyses of concordance lines in Tables 20 to 22 further showed that the applicants used several similar persuasive strategies to highlight their relevant qualifications. For instance, many of them emphasized the rankings of their undergraduate universities, as well as the uniqueness of their previous achievements. For instance, in one concordance line of Table 22, one applicant for a comparative literature program wrote the following sentences as the opening paragraph of her PS:

> Example (a): "As a senior student in one of the China's top universities - X University, which boasts a long history of language education with its distinctive traits and traditional superiority, I am submitting this statement in support of my application for acceptance into your comparative literature Postgraduate Program."
>
> (Comparative Literature 1)

From this excerpt, it can be clearly observed that the applicant strongly proposed that the reputation of her undergraduate university played a significant role in her background. Similar sentences can be found in many other PS samples. If one applicant came from a university that may not be well-known in the western world, then he/she would try to discuss the reputation of his/her undergraduate programs thereby stressing his/her uniqueness. Similar persuasive strategies were also found in concordance lines of the category "Background" in Tables 20 and 21. Applicants of the collected PS-samples

specifically emphasized their rankings in their departments, or the significance of research projects they participated in. The above argument can be supported by the following excerpts from the compiled corpora.

> Example (b): "Undergraduate study provides me with a sound base in animal genetics, animal nutrition, microbiology, biochemistry and physiology. I paid great attention to my kernel courses and maintained an X GPA while handling a full complement of English Test (GRE and TOEFL). Ranking top X% of the class, I achieved the National Scholarship which was only awarded the top, three years of merit student and top level scholarship of X which were all rewarded outstanding students."
>
> (Biology 3)

> Example (c): "For my excellent research performance, I was honoured to participate in the Summit Forum for Urban Water hosted by Chinese Academy of Engineering (CAE) and my Alma Mater."
>
> (Environmental Studies 3)

Meanwhile, the results in Tables 20 to 22 also presented several differences across the compiled corpora. First, the Sciences Corpus included more concordance lines of "study" than other corpora in Table 20. Second, Table 22 presents that the notion of "China/Chinese" was more frequently emphasized in the Arts Division than in the science division. The Arts Corpus had the most concordance lines of "Chinese/China" in Table 22, followed by the Social Sciences Corpus. Concordance lines of the Arts Corpus were mainly in the "Background" category, which described the applicants' Chinese culture-related skills (e.g. Teaching Chinese as a Foreign Language). By comparison, concordance lines of the Social Sciences Corpus mainly discussed Chinese cultural and

social issues and their relations with the applicants' study and research prospects. Typical concordance lines of "China/Chinese" of the Arts Division in Table 22 were as follows:

Example (d): "Because of the courses I have taken in the university, such as Linguistics, Ancient Chinese, Modern Chinese and Chinese Literature, and the experiences of teaching Chinese to students coming from other countries, I was inspired to be an ambassador of sorts of Chinese characters and culture."

(Applied Linguistics 3)

Example (e): "China has developed at a high speed in recent years, and with the globalization waves, the intercultural communication with other countries becomes more and more frequent day by day."

(Communication 3)

Third, similar to the results in Sections 4.2.1 and 4.2.2, the notion of "research" received different attention between applicants in the Arts Division and the Sciences Division. In Table 21, corpora of the Sciences Division had more concordance lines of "research" than corpora in the Arts Division. Further content analyses revealed that applicants in the Sciences Division tended to provide concrete descriptions of their previous research projects and future plans. By contrast, the descriptions of research interests by applicants in the Arts Division tended to be shorter and more general. Such difference can be illustrated by the comparison between the Examples (f) and (g) below:

Example (f): "The curriculum in area of synthetic organic chemistry such like Medicinal Chemistry, Synthetic Methods in Organic Chemistry, Organometallic Chemistry, Organic Spectroscopy and Structure Determination exactly fit my research interests."

(Chemistry 2)

Example (g): “the international relationships among China and other countries are the basis for the research I should do.”

(Public Administration 3)

Finally, results of the E-PS Corpus in Tables 20 to 22 followed the same tendencies of other unedited PS corpora. Its concordance lines of “study” distributed evenly in the “Credential” and “Prospect” categories in Table 20 while it had more concordance lines in the “Credential” category than in the “Prospect” category in Table 21. The distribution of concordance lines of the E-PS Corpus in Table 22 was also similar to other unedited PS corpora.

4.3 Research Question 3: Functional Moves of the Collected CEAL-PSs

The collected CEAL-PSs were coded based on the coding scheme in Table 10, in which a prototypical PS is coded as four general moves (“Introduction-IN,” “Establish Credentials-EC,” “Reasons for Application-RA,” and “Conclusion-CO”), with three to five sub-moves in each category. The functional move analysis results are presented in Tables 23 to 25.

Table 23 presents the frequencies of the major moves and their sub-moves across the compiled corpora. Following the results in Table 23, average length of the general moves in each compiled corpus calculated by NVivo is shown in Table 24. The average length (presented as percentages in Table 24) was calculated as follows: first, the length of general moves in each PS sample was achieved by NVivo as percentages; then, all the percentages of one move were added up and divided to get a final mean value. For instance, the mean value of the IN move for the Arts Corpus in Table 24 was achieved by

adding up all the percentages of IN moves in the Arts Corpus and then dividing 17 (17 samples in the Arts Corpus have the IN move).

Finally, the average lengths of five sub-moves ("Research-RE," "Work Experience-WE," "Other-OT," "Positive Gains-PG," and "Disciplinary/Research Reasons-DR") were also calculated since the results in Table 23 suggested potential inter-corpus differences among these moves. The results of the sub-moves are shown in Tables 24 and 25.

4.3.1 Analysis of Move Frequency

The results of Table 23 suggested that the prototypical move structure in Table 10 was followed by most collected PS samples. The EC and RA moves were the most well-established moves among the four general moves. The EC move appeared in all the collected CEAL-PSs and the RA move was missed by only three PS samples. The other two moves, IN and CO, also appeared in most PS samples. Among the 120 CEAL-PSs, only 14 lacked the IN move and 15 lacked the CO move. Thus, the IERC model (Introduction, Establish Credentials, Reasons for Application, and Conclusions) in Table 10 might be a common move structure for PS writing.

In terms of sub-moves, Table 23 shows several similarities as well as differences across the compiled corpora. The compiled corpora were similar with each other in the most frequent sub-moves of each general move. For instance, the "Background" move was used as the most frequent sub-move in the IN move by most corpora except for the Business Corpus. In the EC and RA moves, the "Education" and "Program/University Attributes" moves were emphasized by all disciplines, as shown by their high frequencies across the compiled corpora. For the CO move, there were slight differences of sub-move frequencies in each corpus, which indicated a lack of clear patterns in this move. On the other hand, inter-corpus variations emerged in some sub-moves, especially in the EC and

RA moves. In the EC move, the sub-move "research" appeared more frequently in corpora in the Sciences Division than those in the Arts Division. By contrast, the sub-moves "Work Experience" and "Other" were more frequently used by corpora in the Arts Division than those in the Sciences Division. Similarly, in the RA move, the sub-move "Disciplinary/research reasons" was frequently emphasized by PSs in the Sciences Division, but not in the Arts Division. The sub-move "Positive gains," by comparison, tended to be more preferred by the Arts Division than the Sciences Division.

Table 23 ***Move Frequency Results of the Compiled Corpora***

Move	The Arts Division			The Sciences Division		E-PS
	Arts	Business	Social Sciences	Engineering	Sciences	
1. Introduction (IN) (106/120)	17	18	20	16	20	15
a) Generalization/attention catcher	8	9	5	10	8	10
b) Background	12	8	10	10	13	11
c) Reasons for application	6	10	11	10	9	7
2. Establishing credentials (EC) (120/120)	20	20	20	20	20	20
a) Education	20	17	15	17	18	17
b) Research	9	11	11	17	20	15
c) Work experience	12	14	10	9	6	11
d) Personal attributes	5	7	6	10	1	5
e) Other experience	13	17	13	8	6	12
3. Reasons for application (RA) (117/120)	20	20	20	19	18	20
a) Background gap	9	9	6	7	3	4
b) Positive gains	11	16	9	10	4	10
c) Program/university attributes	17	12	17	16	17	18
d) Disciplinary/research reasons	9	7	11	13	15	10
4. Conclusion (CO) (105/120)	17	18	19	19	16	16
a) Goals/prediction of future	7	10	11	10	9	7
b) Self-evaluation	8	9	7	13	6	12
c) Final Appeal	13	9	11	5	10	9

Note. "X/120" refers to how many samples in the collected PSs included the target move (e.g. Introduction).

4.3.2 Analysis of Move Length

The Kruskal-Wallis test was applied to the results in Tables 24 and 25 for determining whether there was any statistical significant difference across the compiled corpora. The results in Table 24 indicated both similarities and differences across the compiled corpora in general move length. As shown in Table 24, the compiled corpora had approximately 10% in the IN move, 27% in the RA move, and 7% in the CO move. Such similarities were further supported by the Kruskal-Wallis test ($p>0.05$). However, the Kruskal-Wallis test in Table 24 also indicated that there was statistically significant difference in the EC move ($p=0.027$). Corpora in the Sciences Division tended to have longer EC moves than those in the Arts Division.

The results in Table 25 showed that the Kruskal-Willis test for the RE and PG moves achieve significant values ($p=0.035$ and 0.03), indicating significant differences in the two sub-moves across the compiled corpora. For the EC-Research move, it followed the findings in Table 24: the Sciences Division tended to have longer EC-Research moves than the Arts Division. Nevertheless, the results were reversed in the RA-Positive gains move. Corpora in the Sciences Division had lower percentage values than those in the Arts Division. The results for other sub-moves in Table 25 were similar across the compiled corpora.

Finally, the results of the E-PS Corpus in Tables 23-25 were similar to other unedited PS corpora again. The E-PS results in Table 23 followed the IERC model. In Tables 24 and 25, the average move length of the E-PS Corpus was close to the results of other unedited corpora.

Table 24 ***Move Length Results across the Compiled Corpora***

	The Arts Division			The Sciences Division		E-PS		
Move	Arts	Business	Social Sciences	Engineering	Sciences		Average	*p*
1. Introduction (IN)	10.51% (17)	9.53% (18)	10.47% (20)	9.72% (16)	11.07% (20)	10.15% (15)	10.24%	0.897
2. Establishing credentials (EC)	53.93% (20)	57.48% (20)	49.62% (20)	59.32% (20)	65.57% (20)	56.80% (20)	57.12%	0.027
3. Reasons for application (RA)	29.02% (20)	23.83% (20)	32.50% (20)	27.61% (19)	18.45% (18)	31.73% (20)	27.19%	0.092
4. Conclusion (CO)	7.26% (17)	8.44% (18)	6.87% (19)	6.75% (19)	7.91% (16)	6.46% (16)	7.28%	0.832

Note. Results are presented as X% (Y), in which the figure Y refers to how many samples (of one corpus) contain the specific move.

Table 25 ***Length of Five Sub Moves across the Compiled Corpora***

	The Arts Division			The Sciences Division		E-PS		
Move	Arts	Business	Social Sciences	Engineering	Sciences		Average	*p*
EC-Research	28.36% (9)	17.21% (11)	25.86% (11)	32.01% (17)	36.30% (20)	26.50% (15)	27.7%	0.035
EC-Work Experience	16.05% (12)	21.04% (14)	19.37% (10)	22.03% (9)	14.86% (6)	21.98% (11)	19.22%	0.571
EC-Other	22.68% (13)	20.51% (17)	24.13% (13)	19.43% (8)	18.05% (6)	18.49% (12)	20.55%	0.788
RA-Positive gains	15.04% (11)	11.63% (16)	15.64% (9)	10.73% (10)	6.92% (4)	25.61% (10)	14.26%	0.03
RA-Disciplinary/research reasons	16.16% (9)	14.86% (7)	28.01% (11)	17.17% (13)	11.19% (15)	18.30% (10)	17.62%	0.076

Note. Results are presented as X% (Y), in which the figure Y refers to how many samples (of one corpus) contain the specific move.

Chapter 5 Discussions, Implications, Limitations, and Future Research Directions

5.1 Discussion of Linguistic Complexity Results

The results in Sections 4.1.1 and 4.1.2 showed that there were few differences across the compiled corpora in lexical diversity and readability, indicating similarities among the collected CEAL-PSs in linguistic complexity.

For results from the lexical diversity analysis, most lexicons used in the compiled corpora fell into the Level One category, which includes 1,000 the most frequent lemmas in English. According to Nation and Coxhead (2001), the reference word list of Range provides a good representation of lexical usage in general academic writing. Thus, the findings in Section 4.1.1 suggest that the PS writing is not genre or discipline unique in lexical diversity. It includes many lexicons that are frequently used across academic genres. The compiled corpora comprised of approximately 10% tokens and lemmas in the "Other" category, which may be attributed to the difficult jargon used in the compiled corpora (see Table 26 in Appendix C for examples).

Meanwhile, the results in Tables 11 and 12 did indicate several slight differences between the Arts Division and the Sciences Division. The Sciences Division tended to include fewer tokens/lemmas at Level One but more in the "Other" category than the Arts Division. One possible explanation for such distinction is the preference of using jargon and terminologies between Arts and Sciences Divisions. Students from the Sciences Division (e.g. engineering and formal science majors) usually receive training of using "standard expression" in their undergraduate writing, such as correctly using standard discipline-specific jargon and terminologies. For instance, CEAL-PSs in the Engineering and Sciences corpora included lexicons such as "apoptosis," "metastasis," and

"Photocatalytic," which are difficult to understand without appropriate knowledge background (see Appendix C for more examples). By comparison, terminologies for many Arts disciplines are also common academic lexicons, such as acquisition (in linguistics), elasticity (in Economics), and party (in Political Science). Therefore, CEAL-PSs of the Sciences Division tend to have a narrower readership than those written for the Arts Division due to the jargon barrier of scientific disciplines. However, overall, PSs for both Arts and Sciences Divisions share similarities in lexical diversity.

The results of the E-PS Corpus in Tables 11 and 12 fell between the results of the Arts and Sciences Divisions, which may be attributed to the fact that the E-PS Corpus itself was composed of edited CEAL-PSs from both Arts and Sciences Divisions. It is somewhat surprising to find that edited CEAL-PSs did not include more lexicons at Level Two and Level Three than the unedited ones. There are two possible explanations for this finding. First, the potential distinction between unedited and edited CEAL-PSs in lexical diversity was weakened by the large token quantity of the compiled corpora. Second, the PS editing process might involve only subtle changes of lexicons, which did not change the general lexical diversity in the edited CEAL-PSs.

The readability analysis results followed the general tendencies reflected in the lexical diversity analysis. The compiled corpora's mean scores ranged between 31.32 and 40.28 for the FRE test and between 13.25 and 14.82 for the FKGL test. Both readability scores indicated the level of senior high school or junior post-secondary institution (Kincaid et al., 1975). Such readability levels are appropriate for the PS genre since PSs for graduate school applications are expected to be produced by writters with senior level university education, who are supposed to exhibit advanced English writing ability.

However, some minor distinctions between the Arts Division and the Sciences Division were found in the readability analysis. The CEAL-PSs in the Arts Division seemed to be easier to read than those in the Sciences Division, as reflected in the results of the Arts and Business Corpora. However, such hypothesis was weakened by the fact that the results of the Social Sciences Corpora suggested that it is the most difficult corpus to comprehend. This finding may be attributed to the individual variations of the collected CEAL-PSs. As shown in Tables 13 and 14, all the examined corpora achieved large standard deviation values, which suggested that individual variation played a key role in the collected CEAL-PSs. Therefore, it can be concluded that PSs written for the Arts and Sciences Divisions might differ from each other in readability, but the present analysis results are not able to support such argument due to individual variations in the collected CEAL-PSs.

Finally, results of the E-PS Corpus were similar to those of the unedited corpora, indicating that the potential editing effect is not observed in the readability analysis. The similarity between the edited and unedited CEAL-PSs in readability may be explained by the mechanism of the readability tests. As shown in Figures 1 and 2, both FRE and FKGL tests calculate readability according to the ratios among total syllables, total words, and total sentences. Thus, the potential editing effect in readability might be weakened by the increase of corpus size and ultimately became invisible in the current study.

In sum, both lexical diversity and readability analyses indicate similarities among the collected CEAL-PSs. Most of them are written with frequent lexicons in academic writing and tend to demonstrate senior high school or junior post-secondary readability levels.

5.2 Discussion of Content Focus Results

The results of Sections 4.2.1 to 4.2.3 showed both similarities and differences in content foci across the compiled corpora, which indicated that although common lexicons, narrative framework, persuasive strategies are widely used in PS-writing, the content focus of one particular PS sample is more or less influenced by the discipline of its target graduate program. Meanwhile, the results of the E-PS Corpus were similar to those unedited ones, which question the potential existence of editing effects.

First, the results in Sections 4.2.1 to 4.2.3 provided empirical evidence for defining PS as an academic self-promotional genre, which supports previous arguments in Brown (2004), Ding (2007), and Samraj and Monk (2008). The self-promotional feature of the PS genre was presented by the high frequencies of self-expressions, as shown in Sections 4.2.1. Self-expressions (i.e. "I," "me," and "my") achieved high rankings in all disciplines in Table 15, which suggested that self-promotion is the primary concern of PS authors. Moreover, academic lexicons appeared frequently in Table 15, which supports PS as an academic genre. The results in Section 4.2.2 further confirmed authors' academic credentials are primarily promoted in PS writing. Finally, although self-expressions were not targeted in the concordance analysis, examples in Section 4.2.3 demonstrated that many concordance lines of the selected lexicons were still directly concerned with self-promotion of academic credentials. Overall, the above findings indicate the academic nature of the PS genre, which distinguish it from general job application letters discussed in Bhatia (1993).

Second, Sections 4.2.2 and 4.2.3 identified one general narrative framework and two common persuasive strategies commonly used by the collected CEAL-PSs. Authors of

the collected CEAL-PSs followed a "self-expression + activity + credential + prospect" narrative framework, in which the authors connected their previous credentials (e.g. achievements in study and research) with their prospective graduate programs. In addition, the authors frequently mentioned the ranking information of their undergraduate universities and programs to highlight their credentials. Socio-cultural issues in China were also discussed by applicants as the background information for their future research/career plans. The reasons behind the two common persuasive strategies may be as follows.

(1) Due to the limitation of education resources, there are great disparities of education quality among schools in China from the secondary level, which causes fierce competition among students in examinations. Consequently, students in China are educated to go to the "best" universities since the start of their secondary education (Dello-lacovo, 2009). Such explicit emphasis naturally influences Chinese students' understanding of "qualification." To many Chinese students, the major proof of one's excellence is to attend a university with high ranking. Therefore, it is no wonder to see that university and even department rankings were repeatedly mentioned in the collected CEAL-PSs.

(2) As shown in the examples (d) to (g) in Section 4.2.3, the lexicons "China" and "Chinese" were used as source-cultural background information to justify the authors' motivations of applications. Such explicit emphasis on demographic origin may be attributed to the applicants' awareness of competition from Western applicants. There has been increasing competition for Western graduate school admission in recent years. For many Chinese applicants, their applications may be potentially hindered by the

differences between Western and Chinese higher education systems. Thus, as a compensatory strategy, many Chinese applicants may choose to highlight their Chinese background as a way to show their uniqueness. This strategy was particularly favoured by applicants from the Arts Division in the collected CEAL-PSs, since research topics in the Arts Division can be easily connected with socio-cultural issues in China.

Third, some differences across the compiled corpora were found. In Section 4.2.1, each corpus included several discipline-specific terminologies. Thus, similar to lexical use patterns found in academic articles (e.g. Samraj, 2002, 2005), there are inter-disciplinary differences within the PS genre in the use of academic terminologies. Furthermore, several distinctions between the Arts Division and the Sciences Division were found, especially in content of research experience and prospects. CEAL-PSs in the Sciences Division demonstrated better expertise in describing research experience and prospects than those in the Arts Division. Such finding is probably caused by the distinctions in the academic training of the PS authors. Despite the pursuit of quality education in China, the distinctions between the Arts and Sciences Division from secondary education still play a key role in Chinese students' academic development at the tertiary level. The notion of "research" (especially empirical research) is not sufficiently emphasized by Chinese curriculums in the Arts Division at the secondary level (Dello-Iacovo, 2009). Such limitation is also extended to Chinese curriculums in the Arts Division at the tertiary level. For instance, there is no course requirement for research methodology in English majors' undergraduate curriculums in China, according to the guideline published by the Ministry of Education of China (Chinese Higher Education Committee of Foreign Language Major, 2000). As a result, Chinese students from the Arts Division tend to lack adequate

research experience. Although in the collected CEAL-PSs such weakness was noticed by many applicants from the Arts Division, there is a mismatch between how the applicants describe their research ability and the desired qualifications by universities. By comparison, Chinese applicants from the Sciences Division have more experience in laboratory work and research design. As a result, they were able to provide specific and well organized descriptions of their research experience and prospects in the collected CEAL-PSs.

Finally, the results of the E-PS Corpus were similar to other unedited PS corpora, which is somewhat unexpected. Before staring the data analysis, I expected that the edited CEAL-PSs would have better discussions regarding the applicants' credentials from both study and research aspects. However, the results of the E-PS Corpus tended to be indistinguishable from other unedited PS corpora. Such finding may be attributed to two factors. First, the E-PS Corpus was comprised by PSs written for various disciplines, which inevitably weakened the frequencies of discipline specific terminologies. Perhaps a more unified E-PS Corpus (e.g. PSs all written for the Sciences Division) will be able to reveal more unique factors of edited PSs. Second, although PSs in the E-PS Corpus were edited by EL1 professional writers, their original authors were still Chinese applicants with many years training in the education system with arts and science divisions. Therefore, it is no wonder that the original E-PS authors would use persuasive strategies similar to those in the unedited CEAL-PSs.

In sum, the above results demonstrate that the authors of the collected CEAL-PSs were aware of the primary communicative purpose of the PS genre: self-promotion for getting admission into the desired academic community. The collected CEAL-PSs mainly

focused on expressions regarding the applicants' academic qualifications. In addition, a general narrative framework was widely used by the collected CEAL-PSs, along with two common persuasive strategies: highlighting academic rankings and discussing Chinese socio-cultural issues.

5.3 Discussion of Functional Move Results

The analysis in Sections 4.3.1 and 4.3.2 explored the functional move structure of the collected CEAL-PSs, which followed analyses from previous studies (Ding, 2007; Samraj & Monk, 2008). Overall, the present analysis results supported the IERC model as one common move structure for the PS genre. In addition, sub-moves of the IERC model were studied across the compiled corpora, which further confirmed them as different strategies for realizing the general moves (Bhatia, 1993). In terms of inter-corpus differences, the functional move analysis provided further support for the findings in Section 4.2. There were distinctions between PSs written for the Arts and Sciences Divisions. Finally, the results of the E-PS Corpus fell between the Arts and Sciences Divisions again, which followed the tendency from previous sections.

First, the results in Section 4.3.1 demonstrated that the majority of the collected CEAL-PSs followed the IERC model across disciplines. Thus, it could be argued that the IERC model is one common move structure of the PS genre. Among the 120 collected CEAL-PSs, the EC and RA moves were only missing in three PS samples, showing that the EC and RA moves probably function as obligatory moves for the PS genre. PSs lacking these two moves, as a result, are likely to be regarded as "unqualified" PSs by graduate admission committees. By comparison, the IN and CO moves were missing in a few samples in the compiled corpora, indicating that these two moves might be quasi-

obligatory for the PS genre. However, considering the fact that the majority of the collected CEAL-PSs included all of the four general moves, it can be concluded that the IERC model is a crucial move structure of PS writing.

Second, in Section 4.3.1, the coded sub-moves of each general move were distinguished by various content foci, which can be defined as move strategies. As defined in Bhatia (1993), these sub-moves are used for the realization of general moves. The results in Table 23 showed that the compiled corpora were similar in many of the strategies that they employed in developing the general moves, which in fact demonstrated the authors' knowledge of the PS genre. Thus, the PS authors of the present study must acquire the PS genre through extra-curricular resources. One tentative explanation is PS samples available online to potential applicants. As mentioned in Chapter 3, the current study collected CEAL-PSs from the online bulletin board Gter, which contains many posts regarding PS writing. The authors of the collected CEAL-PSs probably browsed through these online resources before writing their own PSs. Thus, although PSs are an occluded genre, the acquisition of this genre is still possible from self-learning.

Third, the results in Section 4.3.2 indicated some distinctions between PSs written for the Arts and Sciences divisions. In Table 25, PSs written for the Sciences Division showed longer RE moves than those for the Arts Division. By contrast, PSs in the Arts Division had longer PG moves than those in the Sciences Division. These results followed the findings in previous sections. As discussed in Section 4.2.4, research experience and prospects are betted developed in the Sciences Division than in the Arts Division due to more research-oriented courses in the Sciences Division curriculums in

China. However, considering the fact that many Western graduate programs in the Arts Division are also research oriented (e.g. Linguistics, Communication, and Sociology), the lack of research training among Chinese applicants from the Arts Division may put them in a disadvantageous situation when competing with applicants from other countries.

Finally, the results of the E-PS Corpus did not differ from those of unedited PS corpora. On the one hand, such finding supports the IERC model as one common model for the PS genre. The general IERC structure does not vary from discipline to discipline, which is in concordance with previous finding of Samraj and Monk (2008), in which PSs written by EL1 and EAL applicants did not differ from each other. On the other hand, the lack of potential editing effects may be attributed to the small sample size the various disciplines of PSs in the E-PS Corpus.

In sum, the functional move analysis proposed the IERC model as one common move structure for the PS genre. Various sub-moves were used in the collected CEAL-PSs, demonstrating the applicants' ability in adopting different strategies for move realization. The results of the Arts and Sciences Divisions were divided by the discussion of research experience and prospects. The E-PSs followed the IERC structure and shared similar results with unedited CEAL-PSs.

5.4 Summary of Key Findings

The results of the present study offered several valuable insights regarding the genre features of PSs. In line with previous studies of the PS genre, self-expressions and academic credentials were found to be explicitly emphasized at the lexical, phrasal, and discoursal levels in the present study, which provided further evidence for defining PSs as an academic self-promotional genre. One common narrative framework of PS writing

was found in the collocation analysis, in which the PS authors explicitly connected their previous credentials with their prospective graduate programs. In addition, the functional move analysis proposed the IERC model as one common move structure for the PS genre.

Meanwhile, the present study found some unique linguistic features of CEAL-PSs. The results in the lexical diversity analysis suggested that the majority of lexicons used in the collected CEAL-PSs were frequent lexicons in general academic writing. The readability analysis showed that the average readability of the collected CEAL-PSs was at senior high-school or junior post-secondary levels. The concordance analysis also identified two common persuasion strategies used in the collected CEAL-PSs: (1) the emphasis of academic rankings and (2) the discussion of Chinese socio-cultural issues.

Furthermore, the results of the present study indicated differences between PSs written for the Arts and Sciences Divisions. PSs written for the Sciences Division tended to include more discipline specific terminologies than those written for the Arts Division. The Sciences Division also had better discussions of research experience and prospects than the Arts Division. The above results can be attributed to the different academic training received in the two divisions.

Finally, although the present study did not find significant distinctions between the edited and unedited CEAL-PSs, it did not reject the existence of the editing effects. As discussed in previous sections, the various disciplines in the E-PS Corpus may hinder the presence of edited information. Thus, in future studies, the potential editing effect should still be explored through more detailed content analysis of edited and unedited CEAL-PSs in the same discipline.

5.5 Theoretical Implications

From a theoretical perspective, the present study establishes the IERC model of the PS genre, which promotes the advancement of genre theories. The discussion of general move structures of different genres has been prevalent in genre research since the development of the IMRC model for research papers (Swales, 1990). Researchers working on move analysis have two general directions: (1) the investigation of sub-moves of well-established genres, such as the various discussions regarding the Introduction section in research papers (e.g. Swales, 1981, 1990); and (2) the exploration of move structures of less studied genres, such as the studies conducted on other promotional genres (e.g. Fortanet, 2008; Hyon, 2008; Loudermilk, 2007). The present study followed the second direction to discuss a less-studied student genre: Personal Statements. Although the PS genre has been defined as academic self-promotional genre in previous literature (Bhatia, 1993), few studies have investigated its genre features (Brown, 2004; Ding, 2007; Samraj & Monk, 2008). The validity of these studies is questionable due to their limited PS samples. By comparison, analyses of the present study were based on a large collection of CEAL-PSs (N=120). The results of the present study not only showed that the IERC model was followed by more than 85% of the collected CEAL-PSs, but also demonstrated how the communicative purpose of the PS genre (self-promotion for desired academic communities) was realized at the lexical, phrasal, and discourse levels. Therefore, based on the findings of the present study, the PS genre can be convincingly defined as an academic promotional genre, and it follows the IERC model as one general move structure.

5.6 Methodological Implications

From a methodological perspective, the present study provides valuable implications for corpus assisted genre analysis. On the one hand, it demonstrated the advantages of using corpus-driven methods in genre analysis. The validity of traditional genre analysis methods (e.g. hand-tagging) has been questioned in recent years due to the concern of their inability of processing large size data (Flowerdew, 2005). By comparison, the corpus-driven methods are able to handle large-sized data (120 samples in the present study) from both quantitative and qualitative perspectives. They increase research validity through data triangulation while also keeping the data analysis process in a timely manner. Considering the increasing availability of language corpora online and the advancement of text technologies (e.g. new scanning and speech recognition techniques), corpus-driven methods will be adopted by more genre studies in the near future.

5.7 Pedagogical Implications

From a pedagogical perspective, the present study highlights the potential benefits of embedding the instruction of the PS genre into the current EAL curriculum in China. Although the results of the present study showed that the authors of the collected CEAL-PSs were able to use the PS genre without formal PS writing instructions, the PS genre is still a valuable genre for EAL instruction in China. First, it is necessary to provide more information of PS writing for Chinese undergraduates due to the significance of PS writing in one's academic career. The searching for well-written PS samples online is a difficult process. The occluded feature makes most PSs written by successful applicants unavailable for novice PS writers. Inappropriate PSs might hurt one's academic pursuits. Therefore, for graduate school applicants, the instruction of the PS genre can assist their

graduate application processes, which might further benefit their long-term academic development.

Second, the results showed that the PS genre shares many similar features with formal academic genres (especially in linguistic complexity and content foci). The instruction of it can also benefit Chinese undergraduates' EAL writing development. As shown in the present study, a well-written PS tends to follow a certain promotional structure and include many standard disciplinary terminologies and academic expressions. Therefore, a good knowledge of the PS genre for one discipline may be transferred to the writing of other types of academic genres (e.g. research papers, and funding proposals). In addition, even for students without graduate study plans, the instruction of the PS genre can be beneficial as well. As shown in the present study, PSs and job application letters are both promotional genres. Currently, job application letters are not explicitly taught in China at the tertiary level. Thus, the instruction of the PS genre can function as an introduction to promotional genres, which will benefit Chinese undergraduates from both academic and professional perspectives.

Third, a better understanding of the PS genre can be helpful for graduate admission committee members as well. As the target audience of the PS genre, graduate admission committee members face the task of reading hundreds of PSs every year. Many graduate admission works in Canada are solely based on paperwork and therefore, one applicant's success of graduate application is often determined by whether his/her PS can be correctly interpreted by graduate admission committee members. Thus, a better understanding of the PS genre can improve graduate admission committee members'

awareness of how Chinese EAL writers approach PS writing, thereby assisting their profile review processes and benefiting Chinese EAL graduate applicants.

Fourth, the present study further supports the ubiquity of the promotional feature in contemporary academic genres. As shown in the present study, the "promotional writing" can be clearly observed in the PS genre, a pre-graduate school academic writing. As suggested by previous studies (e.g. Swales, 1981, 1990), "promotional writing" is crucial for certain sections of academic genres. Thus, learning the PS genre can be significant for prospective graduate students since it can function as an implicit introduction of the promotional feature in future formal academic genres. From this perspective, the PS might be regarded as a bridging genre which transforms undergraduates into formal academic communities.

5.8 Limitations and Future Research Directions

There are five limitations of the present study, which need to be considered for the development of follow-up studies in the future. First, the present study's results indicated some restrictions of current corpus linguistics methods. The depth of content analysis was limited in the present study by the large size of data. Although the application of NVivo can improve the coding efficiency, the coding procedure of functional move analysis may still be influenced by researcher's bias. Some recent studies have attempted to solve the problem by using customized corpus analysis programs (e.g. Biber et al., 2004). However, the challenge of sentence auto coding will still affect corpus-driven genre analysis for a long time. Thus, although the present study indicates some promising prospects of corpus-driven methods, the traditional handing tagging method is still valid at this point and it may be more appropriate for thorough analyses of single moves of the PS genre.

Coding reliability is another issue for consideration. The present study used an intra-coder reliability to check the date coding results. Future studies can adopt inter-coder reliability to improve the validity of the move coding process.

Second, the majority of CEAL-PSs in the present study were collected from current graduate school applicants. Whether these CEAL-PSs have helped their authors to get admissions into their desired graduate programs is unknown. As a result, the generalizability of the present study is limited. An investigation of successful CEAL-PSs would be a good research design for future studies of the PS genre. However, how to get sufficient successful CEAL-PSs would be a great challenge, due to the privacy regulations in many Western institutions as discussed in Chapter 2. Therefore, future studies could adopt the analysis procedure in the present study, with more emphasis on potential differences between unsuccessful and successful PSs.

Third, the present study did not investigate the sub-moves in the IERC model. As shown in many previous studies regarding the introduction section in research papers (e.g. Swales, 1981; 1990), detailed content and structural analysis of one general move of a genre may reveal some fundamental features of that genre. Therefore, the inter-disciplinary variations can be further studied through analysing sub moves of each general move of the PS genre (e.g. the detailed analysis of the EC move or the RA move).

Fourth, unlike Brown (2004) and Samraj and Monk (2008), the present study was not able to include interviews with members of admission committees, whose "insider views" can provide valuable information regarding expected features in PS writing. Therefore, future studies can also focus on the expectations of admission committee members in the

PS genre. Admission committee members are the target readers of the PS genre but their views have not been sufficiently investigated so far.

Fifth, due to the limitation of data size, the present study did not find significant difference between unedited and edited CEAL-PSs. However, the comparison between unedited and edited CEAL-PSs is still a topic worth investigation because it is crucial to know whether the booming online editing services threat academic integrity of the graduate admission processes. Future studies can be conducted on a larger edited PS sample size or focus on the comparison of unedited and edited PS samples in one discipline.

5.9 Conclusion

The present study conducted multi-perspective analyses on CEAL-PSs, which explored the PS genre's linguistic complexity, content foci, and functional move structure. According to my knowledge, this study is the first in applied linguistics literature. Thus, the findings of the present study can provide significant implications for academic promotional genre theories and research methods. The study also has valuable pedagogical implications for EAL writing instruction in China.

In general, the present study defines PS as an academic self-promotional genre following the IERC model ("Introduction," "Establish Credentials," "Reasons for Application," and "Conclusion") as one common move structure. The application of corpus-driven analysis methods is also significant as it indicates the advantages of applying corpus linguistics in genre analysis. In addition, the present study highlights the pedagogical values of the PS genre. Due to its bridging functions in undergraduates'

academic as well as professional developments, it can be effectively incorporated into ESP/EAP teaching in China.

Moreover, the differences between PSs written for the Arts and Sciences Divisions have some tentative implications for the education reform in China. Although China has pursued "*Shu-Zhi-Jiao-Yu*" (quality education) in Chinese education system reform for decades (Dello-Iacovo, 2008), the results in the present study clearly showed that the gap between the Arts and Sciences Divisions is still deep. Thus, for Chinese undergraduates' comprehensive development, some curriculum reforms are needed in China, especially in the area of research methodology education in the Arts Division.

Finally, the present study also reveals the insufficiency of research in the PS genre. PSs play important roles in undergraduates' academic development whereas its genre features and potential instruction in the university context are still neglected by genre scholars. Therefore, there is an urgent need for more studies on the PS genre from multiple perspectives in the future.

References

Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora, 4*(1), 71-83.

Anthony, L. (2009a). Issues in the design and development of software tools for corpus studies: the case for collaboration. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 87-104). London, UK: Continuum.

Anthony, L. (2009b). AntWordProfiler 1.21. Retrieved from: http://www.antlab.sci.waseda.ac.jp/software/awp121/readme_file_antwordprofiler1.21.txt

Anthony, L. (2011). AntConc build 3.3.0. Retrieved from: http://www.antlab.sci.waseda.ac.jp/software/antconc330/AntConc_readme.pdf

Bakhtin, M. L. (1952–1953 [1986]). *Speech genres and other late essays*. V. W. McGee (trans.) and C. E. Emerson & M. Holquist (Eds.). Austin, TX: University of Austin Press.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Baker, P. (Ed.) (2009). *Contemporary corpus linguistics*. London: Continuum.

Berber, S. T. (2005). The influence of reference corpus size on WordSmith tools keywords extraction. *The ESP Specialist, 26*(2), 183-204.

Bhatia, V. (1993). *Analyzing genre: Language use in professional settings*. London, UK: Longman.

Biber, D., Csomay, E., Jones, J. J., & Keck, C. (2004). A corpus linguistics investigation of vocabulary-based discourse units in university registers. In U. Connor & T. A.

Upton (Eds.). *Applied corpus linguistics: A multidimensional perspective* (pp. 53-72). Amsterdam, NL: Editions Rodopi B. V.

Brown, R. M. (2004). Self-composed rhetoric in psychology personal statements. *Written Communication, 21*(3), 242-260.

Chujo, K., Utiyama, M., & Miura, S. (2006). Using a Japanese-English parallel corpus for teaching English vocabulary to beginning-level students. *English Corpus Studies, 13*, 153-172.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Dello-Iacovo, B. (2009). Curriculum reform and 'Quality education' in china: An overview. *International Journal of Educational Development, 29*(3), 241-249.

Ding, H. (2007). Genre analysis of personal statements: Analysis of moves in application essays to medical and dental schools. *English for Specific Purposes, 26*(3), 368-392.

Diniz, L (2005). Comparative review: TextSTAT 2.5, AntConc 3.0, and Compleat Lexical Tutor 4.0. *Language, Learning and Technology, 9*(3), 22-27.

Dudley-Evans, T. (1994). Genre analysis: An approach to text analysis for ESP. In M. Coulthard (Ed.). Advances in written text analysis (pp. 219-228). London, England: Routledge.

Corbett, J. (2006). Genre and genre analysis. In K. Brown (Ed.), *Encyclopaedia of language and linguistics* (2nd Ed.) (pp. 26-32). Oxford: Elsevier.

Feton, A. (2006). *Weft QDA User Manual*. Retrieved from http://www.pressure.to/qda/doc/wefthelp.pdf

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology 32*, 221–233.

Flowerdew, J. (1990). English for specific purposes-A selective review of the literature. *ELT Journal, 44*(4), 326-337.

Flowerdew, L. (1998). Corpus linguistic techniques applied to textlinguistics. *System, 26*(4), 541-552.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes, 24*(3), 321-332.

Gter (n.d.) Retrieved from http://www.gter.net/

Halliday, M. A. K. & Hasan, R. (1989). *Language, context and text: aspects of language in a social-semiotic perspective* (2nd Ed.). Oxford, UK: Oxford University Press.

Heatley, A., Nation, I.S.P. & Coxhead, A. (2002). RANGE and FREQUENCY programs. Retrieved from http://www.vuw.ac.nz/lals/staff/Paul_Nation.

Henry, A., & Roseberry, R. L. (2001). A narrow-angled corpus analysis of moves and strategies of the genre: 'Letter of application. *English for Specific Purposes, 20*(2), 153-167.

Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes, 7*(2), 113-121.

Huang, L.-S. (2010). The potential influence of L1 (Chinese) on L2 (English) communication. *ELT Journal, 64*(2), 115-164.

Huang, L.-S. (2011). Corpus-aided language learning. *ELT Journal, 65*(4), 481-484.

Hunt, K. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research report No. 3. Champaign, IL: NCTE.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.

Hunston, S. (2006). Corpus Linguistics, In K. Brown (Ed.) *Encyclopaedia of language and linguistics* (2nd Ed.) (pp. 234-248), Oxford, UK: Elsevier.

Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam, NL: John Benjamins.

Hyland, K. (2003). Dissertation acknowledgements: The anatomy of a Cinderella genre. *Written Communication, 20*(3), 242-268.

Hyland, K. (2009). *Academic discourse: English in a global context*. NY: Continuum.

Hyland, K., & Tse, P. (2009). The leading journal in its field': Evaluation in journal descriptions. *Discourse Studies, 11*(6), 703-720.

Hyon S (1996). Genre in three traditions: implications for ESL. *TESOL Quarterly 30*(4), 693–722.

Hyon, S. (2008). Convention and inventiveness in an occluded academic genre: A case study of retention–promotion–tenure reports. *English for Specific Purposes, 27*(2), 175-192.

IBM Corp. (2010). IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY: IBM Corp.

Kincaid, J. P., Fishburne, R. P. Jr., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for Navy enlisted personnel. *Research Branch Report (pp.*

8-75). Millington, TN: Naval Technical Training, U. S. Naval Air Station, Memphis, TN.

Kwary, D. A. (2011). A hybrid method for determining technical vocabulary. *System, 39*(2), 175-185.

Labov, W. (1966). *The social stratification of English in New York City*. Washington, D. C.: Centre for Applied Linguistics.

Loudermilk, B. C. (2007). Occluded academic genres: An analysis of the MBA thought essay. *Journal of English for Academic Purposes, 6*(3), 190-205.

Maier, P. (1992). Politeness strategies in business letters by native and non-native English speakers. *English for Specific Purposes, 11*, 189-205.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech 70*(2), 151–167.

Morrow, P. R. (1989). Conjunct use in business news stories and academic journal articles: A comparative study. *English for Specific Purposes, 8*(3), 239-254.

Nation, I. S. P. (2004) A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.) *Vocabulary in a second language: Selection, acquisition and testing* (pp. 3-13). Amsterdam: John Benjamins.

Newby, H. (1977). In the field: Reflections on the study of Suffolk Farm Workers. In C. Bell & H. Newby (Eds.). *Doing sociological research*. London: Allen and Unwin.

QSR International Pty. (2008). NVivo qualitative data analysis software, Version 8.

Reppen, R., Ide, N., & Suderman, K. (2005). *American National Corpus (ANC) second release*. Philadelphia, PA: Linguistic Data Consortium.

Salager-Meyer, F. (1990). Discoursal flaws in medical English abstracts: A genre analysis per research- and text-type. *Text, 10*(4), 365-384.

Salager-Meyer, F., Defives, G., Jensen, C., & De Filipis, M. (1989). Principal component analysis and medical English discourse: An investigation into genre analysis. *System, 17*(1), 21-34.

Samraj, B. (2002). Introductions in research articles: variations across disciplines. *English for Specific Purposes*, *21*(1), 1–17.

Samraj, B. (2005). An exploration of a genre set: research article abstracts and introductions in two disciplines. *English for Specific Purposes*, *24*(2), 141–156.

Samraj, B., & Monk, L. (2008). The statement of purpose in graduate program applications: Genre structure and disciplinary variation. *English for Specific Purposes, 27*(2), 193-211.

Saville-Troike, M. (1982). *The ethnography of communication*. Oxford, UK: Basil Blackwell.

Scott, M. (2001). Mapping key words to problem and solution. In M. Scott & G. Thompson (Eds.). *Patters of text: In honour of Michael Hoey* (pp. 109-127). Amsterdam, NL: John Benjamins.

Scott, M., (2008). WordSmith Tools version 5, Liverpool: Lexical Analysis Software.

Swales, J. M. (1981). *Aspects of article introductions*. Birmingham, UK: The University of Aston, Language Studies Unit.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.

Swales, J. M. (1996). Occluded genres in the academy: The case of the submission letter. In E. Ventola & A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp. 45–58). Amsterdam, NL: John Benjamins.

Swales, J. M. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Ed.). *Academic discourse* (pp. 150–164). London, UK: Longman, Pearson Education.

The British National Corpus, version 3 (BNC XML Edition). (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from http://www.natcorp.ox.ac.uk/

Thompson, G. (2004). *Introducing Functional Grammar* (2nd ed.). London, UK: Arnold.

Upton, T. A., & Connor, U. (2001). Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre. *English for Specific Purposes, 20*(4), 313-329.

Wang, K. & Nation, I. S. P. (2004). Word meaning in academic English: Homography in the Academic Word List. *Applied Linguistics 25*(3), 291-314.

Widdowson, H. G. (1998). Context, community and authentic language. *TESOL Quarterly*, *32*(4), 705–716.

Widdowson, H.G. (2002). Corpora and language teaching tomorrow. Keynote lecture delivered at 5th teaching and language corpora conference, Bertinoro, Italy, 29 July.

Xu, J. and Jia, Y. (2009). Readability Analyser 1.0: A text difficulty analyzing tool. Beijing: The National Research Centre for Foreign Language Education, Beijing Foreign Studies University.

Appendix A

Invitation of research participation

Genre Features of Personal Statements by Chinese English-as-an-Additional-Language Writers: A Corpus-Driven Study

You are invited to participate in a study entitled "**Genre Features of Personal Statements by Chinese English-as-an-Additional-Language Writers: A Corpus-Driven Study**" that is being conducted by X. X is a GRADUATE STUDENT in the department of Linguistics at the University of Victoria and you may contact him if you have further questions by e-mail. As a GRADUATE student, I am required to conduct research as part of the requirements for a degree in Applied Linguistics. It is being conducted under the supervision of Dr. X. You may contact my supervisor at XXX-XXX-XXXX.

Purpose and Objectives

The purpose of this research project is to examine the linguistic features and writing styles of statements of research interests written by Chinese-speaking, graduate-level, EAL (English as an Additional Language) students.

Importance of this Research

The research has significant pedagogical implications. The results of it can provide EAL teachers a better understanding of the linguistics features and writing styles of the statement of research interests.

Participants Selection

You are being asked to participate in this study because you are Chinese-speaking graduate student who are current studying in an English-speaking university. In addition, please answer the following questions to provide me some language background information of you.

1) When you apply for graduate schools, have you passed the minimum requirements of the schools you applied?
2) Could you please tell me your latest TOEFL or IELTS score (you can give me the exact number, or a general range like 80+, 100+ and so on)?

What is Involved

If you agree to voluntarily participate in this research, your participation will include providing your statements\ written for graduate school applications.

Inconvenience

There are no known or anticipated inconveniences to you by participating in this research.

Risks

There are no known or anticipated risks to you by participating in this research.

Benefits

The potential benefits of your participation in this research include improving your understanding of your own academic writing; helping to advance the study of EAL writing. **Also, I will provide feedback on your writing sample to benefit your writing skills.**

Voluntary Participation

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation. If you do withdraw from the study your data will be deleted from the researcher's computer right away.

Anonymity

In terms of protecting your anonymity, all participants' ids will be removed during data analysis, and their names will be replaced with pseudonyms in the research report. **ALSO, ANY PERSONAL INFORMATION APPEARING IN THE COLLECTED WRITING SAMPLES WILL BE DELETED BEFORE ANALYSIS.**

Confidentiality

Your confidentiality and the confidentiality of the data will be protected by a password only known by the researcher.

Dissemination of Results

It is anticipated that the results of this study will be shared with others in the following ways: 1. Presentation at scholar meetings; 2. Publication of journal paper. 3. M. A. Thesis

Disposal of Data

Data from this study will be disposed 2 years after completion of the study. The data will be deleted completely from the researcher's computer. **THE RESULTS OF THE ANALYSIS WILL BE DISSEMINATED VIA THE INTERNET AS ALL UVIC THESES ARE PUBLISHED ON D-SPACE.**

Contacts

Individuals that may be contacted regarding this study include the researcher. In addition, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the Human Research Ethics Office at the University of Victoria.

By completing and submitting the questionnaire, **YOUR FREE AND INFORMED CONSENT IS IMPLIED** and indicates that you understand the above conditions of participation in this study and that you have had the opportunity to have your questions answered by the researchers.

Please retain a copy of this letter for your reference.

Appendix B

Websites for the Data Collection of PSs by EL1 Speakers

Abroad Essay: http://www.abroadessay.com/chinese/index.asp
51 Editing: http://www.51editing.com/
Shine Writer: http://www.shinewrite.com/download/s1/ps.pdf
PS Writer: http://www.pswriter.com
Liuxue Today: http://www.liuxuetoday.com
Essay Edge: http://www.essayedge.com/promo/samplework#grad
Help You Apply: http://www.helpyouapply.org/writing.jsp
Accepted: http://www.accepted.com/grad/sampleessays.aspx

Appendix C

Table ***26 Sample Complex Terminologies in the Collected PS Samples***

Apoptosis	Mesoporous	Peroxide	Vehement
Artemisinin	Metallographic	Photocatalytic	
Bromophenylacetic	Metastasis	Radiotherapy	
Calorimetry	miRNA	Thermoelectric	
Electromagnetics	Organometallic	Thermodynamic	

Appendix D

Lexical keyness of the collected PS samples

The lexical keyness analysis compares the word frequency list of a target corpus A with the word frequency list of a reference corpus B, in order to determine which words occurs statistically more frequently in word list A than in word list B (Baker, 2006). Compared with the lexical frequency analysis, the lexical keyness analysis provides more reliable data regarding the key lexicons of target corpora.

The current analysis employed the British Academic Written English (BAWE) corpus as the reference corpus for the lexical keyness analysis. This corpus contains 2761 pieces of proficient assessed student writing (6,506,995 words in total), which are fairly evenly distributed across four broad disciplinary areas (Arts, Social Sciences, Life Sciences and Physical Sciences).

Table 27 ***The Lexical Keyness Analysis Results of the Complied Corpora***

	The Arts Division						The Sciences Division				E-PS	
	Arts	Freq.	Business	Freq.	Social Sciences	Freq.	Engineering	Freq.	Sciences	Freq.	E-PS	Freq.
1	my	2814	I	2647	I	2168	I	2521	my	2552	my	2256
2	I	2398	my	2371	my	1994	my	2434	I	2399	I	2130
3	Chinese	1102	me	833	me	910	engineering	662	me	724	me	764
4	me	838	myself	274	china	423	me	593	research	454	university	250
5	teaching	464	statistics	236	media	326	university	336	engineering	383	research	219
6	English	403	am	224	Chinese	257	graduate	288	graduate	308	am	207
7	culture	391	courses	216	economics	219	research	287	university	274	professor	193
8	language	338	internship	215	communication	219	am	273	am	239	school	176
9	arts	284	university	213	am	213	internship	216	undergraduate	227	program	141
10	academic	284	estate	197	university	186	logistics	206	study	204	china	139
11	literature	279	China	190	research	186	knowledge	143	program	187	career	137
12	China	266	knowledge	170	school	170	undergraduate	137	computer	180	internship	136
13	translation	265	program	159	psychology	169	electrical	135	science	179	undergraduate	130
14	university	229	graduate	151	myself	166	myself	126	courses	158	engineering	126
15	courses	222	career	146	courses	163	courses	119	physics	158	accounting	125

Table 28 ***Collocates of "Study" in the Compiled Corpora***

	Arts	Freq.	Business	Freq.	Engineering	Freq.	Sciences	Freq.	Social Sciences	Freq.	E-PS	Freq.
1	my	31	I	32	my	28	my	43	I	27	I	20
2	I	25	my	27	I	25	I	34	my	24	my	18
3	further	8	graduate	10	graduate	8	research	10	me	11	me	17
4	undergraduate	7	further	7	years	5	me	10	further	11	year	16
5	graduate	6	statistics	5	year	5	graduate	9	economics	10	further	16
6	Chinese	5	me	5	undergraduate	5	further	9	undergraduate	6	undergraduate	13
7	working	4	research	4	want	4	undergraduate	6	graduate	6	experience	13
8	translation	4	year	3	university	4	field	5	work	4	believe	6
9	me	4	social	3	research	4	year	4	university	3	your	5
10	language	4	perhaps	3	more	4	physics	4	research	3	years	5

Table 29 ***Collocates of "Research" in the Compiled Corpora***

	Arts	Freq.	Business	Freq.	Engineering	Freq.	Sciences	Freq.	Social Sciences	Freq.	E-PS	Freq.
1	I	27	I	16	I	35	my	43	my	31	my	34
2	my	24	my	18	my	31	I	36	I	23	I	33
3	me	7	study	4	work	9	me	14	me	9	me	8
4	interest	6	social	4	scientific	8	study	10	project	5	work	7
5	abilities	5	media	4	me	8	do	10	many	5	project	6
6	work	4	year	3	experience	7	doing	9	experience	5	skills	4
7	well	4	team	3	projects	5	scientific	8	Professor	4	related	4
8	have	4	myself	3	interests	5	future	7	interest	4	background	4
9	field	4	financial	3	found	5	program	6	study	3	undergraduate	3
10	conduct	4	experience	3	doing	5	area	6	some	3	professor	3