A CONTRASTIVE STUDY OF UNIVERSITY ENGLISH-MEDIUM INSTRUCTION LECTURES GIVEN BY NATIVE AND NON-NATIVE ENGLISH LECTURERS: RHETORICAL MOVE STRUCTURE AND FORMULAIC LANGUAGE

Lin Deng

A Thesis Submitted in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy in English Language Studies

Suranaree University of Technology

Academic Year 2019

การศึกษาเปรียบเทียบการบรรยายในระดับมหาวิทยาลัยที่ใช้ภาษาอังกฤษเป็นสื่อโดยผู้บรรยายที่เป็นเจ้าของภาษาและที่ไม่ใช่เจ้าของภาษา: โครงสร้างอัตถภาคและวัจนะสูตรสำเร็จ
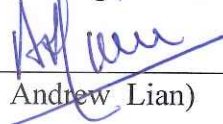
นายหลิน เติ้ง

A CONTRASTIVE STUDY OF UNIVERSITY ENGLISH-MEDIUM INSTRUCTION LECTURES GIVEN BY NATIVE AND NON-NATIVE ENGLISH LECTURERS: RHETORICAL MOVE STRUCTURE AND FORMULAIC LANGUAGE
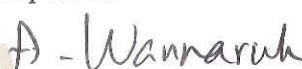
Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

_____

(Prof. Dr. Andrew Lian)

Chairperson

_____

(Assoc. Prof. Dr. Anchalee Wannaruk)

Member (Thesis Advisor)

_____

(Asst. Prof. Dr. Passapong Sripicharn)

Member

_____

(Asst. Prof. Dr. Darunee Dujsik)

Member

_____

(Asst. Prof. Dr. Songphan Choemprayong)

Member

_____   _____

(Assoc. Prof. Flt. Lt. Dr. Kontorn Chamniprasart)   (Asst. Prof. Dr. Thara Angskun)

Vice Rector for Academic Affairs   Dean of Institute of Social Technology

and Internationalization

หลิน เติ้ง : การศึกษาเปรียบเทียบการบรรยายในระดับมหาวิทยาลัยที่ใช้ภาษาอังกฤษ
เป็นสื่อโดยผู้บรรยายที่เป็นเจ้าของภาษาและที่ไม่ใช่เจ้าของภาษา: โครงสร้างอัตถภาคและ
วัจนะสูตรสำเร็จ (A CONTRASTIVE STUDY OF UNIVERSITY ENGLISH-MEDIUM
INSTRUCTION LECTURES GIVEN BY NATIVE AND NON-NATIVE ENGLISH
LECTURERS: RHETORICAL MOVE STRUCTURE AND FORMULAIC
LANGUAGE) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.อัญชลี วรรณรักษ์, 342 หน้า

การเรียนการสอนที่ใช้ภาษาอังกฤษเป็นสื่อ (English Medium Instruction, EMI) ใน
มหาวิทยาลัยในประเทศจีนที่ยังอยู่ในระดับที่ไม่ดีนักเป็นแรงจูงใจในการทำวิจัยเรื่องนี้ เพื่อศึกษา
เปรียบเทียบโครงสร้างอัตถภาคและวัจนะสูตรสำเร็จที่ใช้ในการบรรยายที่ใช้ภาษาอังกฤษเป็นสื่อ
โดยผู้บรรยายที่เป็นเจ้าของภาษาและที่ไม่ใช่เจ้าของภาษา

งานวิจัยเรื่องนี้ใช้คลังข้อมูลในการวิเคราะห์สองกลุ่มคือคลังข้อมูลผู้บรรยายที่เป็นเจ้าของ
ภาษา (Corpus of native English Lecturers, CEL) ประกอบด้วย การบรรยายซึ่งมาจากคลังข้อมูล
MICASE และ BASE จำนวน 12 รายการ และคลังข้อมูลผู้บรรยายชาวจีน (Corpus of native Chinese
Lecturers, CCL) ประกอบด้วยการบรรยายของผู้บรรยายชาวจีนระดับมหาวิทยาลัยในประเทศจีน
จำนวน 12 รายการ กรอบการวิเคราะห์การบรรยาย (lecture genre model) ดัดแปลงจากงานวิจัยของ
Lee's (2016) ส่วนการวิเคราะห์โครงสร้างอัตถภาค (Move) และอนุวัจน์ (Step) ของคลังข้อมูล
ทั้งสองยึดกรอบการวิเคราะห์ของ Swales ผลการวิจัยพบว่าคลังข้อมูลทั้งสองประกอบด้วย 12
อัตถภาค และโครงสร้างอัตถภาคของทั้งสองคลังข้อมูลมีความคล้ายคลึงกัน แต่พบว่ามีความต่างกัน
ของอัตถภาคและอนุวัจน์ในเชิงของ Theme Network Building จากมุมมองด้านภาษาอังกฤษ
ในฐานะภาษากลาง พบอัตถภาคแบบผสมในรูปแบบของการเกิดซ้ำของอัตถภาคในการบรรยายจาก
คลังข้อมูลผู้บรรยายที่เป็นชาวจีน

ในส่วนของวัจนะสูตรสำเร็จ พบประเภทและจำนวนของวัจนะสูตรสำเร็จจากคลังข้อมูล
ผู้บรรยายที่เป็นชาวจีนน้อยกว่าในคลังข้อมูลของผู้บรรยายที่เป็นเจ้าของภาษา ตามโครงสร้าง
สามารถแบ่งวัจนะสูตรสำเร็จได้เป็น 8 กลุ่มตามแนวคิดของ Biber et al. (1999) and Wang (2017)
ประกอบไปด้วย NP fragment, PP fragment, AP fragment as modifier, AP fragment as complement,
Clausal fragment, VP fragment, AdP fragment, และ ConjP fragment ผลวิจัยพบว่าการบรรยายจาก
ทั้งสองคลังข้อมูลมีลักษณะผสมระหว่างการพูดและการเขียน ซึ่งสอดคล้องกับงานวิจัยของ Biber
et al. (2004) จากการวิเคราะห์วัจนะสูตรสำเร็จเฉพาะที่พบในกลุ่มผู้บรรยายชาวจีนไม่พบว่ามีปัญหา
ที่เกี่ยวกับการสื่อสารเกิดขึ้นในระหว่างการบรรยายในห้องเรียนที่ใช้ภาษาอังกฤษเป็นสื่อ

ผลงานวิจัยยังพบว่าวัจนะสูตรสำเร็จมีความสัมพันธ์กับจุดประสงค์ในการสื่อสารของ
อัตถภาคและอนุวัจน์ทั้งทางตรงและทางอ้อม จากการใช้ในสถานการณ์จริงผู้บรรยายชาวจีนใช้
วัจนะสูตรสำเร็จประเภทที่บ่งบอกอัตถภาคน้อยกว่ากลุ่มผู้บรรยายที่เป็นเจ้าของภาษา

ผลงานวิจัยทั้งในด้านโครงสร้างอัตถภาคและวัจนะสูตรสำเร็จมีส่วนส่งเสริมและ
พัฒนาการเรียนการสอนของผู้บรรยายชาวจีนที่ใช้ภาษาอังกฤษเป็นสื่อโดยเฉพาะกลุ่มครูฝึกสอน
และครูที่มีประสบการณ์ในการสอนน้อย

สาขาวิชาภาษาต่างประเทศ
ปีการศึกษา 2562

ลายมือชื่อนักศึกษา _Lin Deng_
ลายมือชื่ออาจารย์ที่ปรึกษา _A. Wannaruk_

LIN DENG : A CONTRASTIVE STUDY OF UNIVERSITY ENGLISH-
MEDIUM INSTRUCTION LECTURES GIVEN BY NATIVE AND NON-
NATIVE ENGLISH LECTURERS: RHETORICAL MOVE STRUCTURE
AND FORMULAIC LANGUAGE THESIS ADVISOR : ASSOC. PROF.
ANCHALEE WANNARUK, Ph.D., 342 PP.

ENGLISH MEDIUM INSTRUCTION LECTURES/ RHETORICAL MOVE
STRUCTURE/ FORMULAIC SEQUENCES/ MOVE-FORMULAICITY
CONNECTION

Motivated by suboptimal language practices in university English Medium
Instruction (EMI) lectures in China, this study compares the rhetorical move structure
and formulaic sequences in EMI lectures given by native English and native Chinese
lecturers.

Twelve lectures were selected from MICASE and the BASE corpus to form Corpus
of native English Lecturers (CEL), and twelve lectures were collected at a Chinese
university to build Corpus of native Chinese Lecturers (CCL). Under Swalesian genre
analysis framework, the present study adapted Lee's (2016) lecture genre model and
identified twelve moves with various steps in CCL and CEL. Move/step distribution
and conventionality results indicate similar rhetorical move structure in CCL and CEL.
However, differences do exist, particularly in moves/steps in the Theme Network
Building Phase. An ELF perspective of CCL lectures reveals move hybridity in the form
of move cycles.

In terms of formulaic sequences (FSs), the study identified significantly less FS
types and tokens by the Chinese EMI lecturers than their native English counterparts.

Structurally, the FSs were classified into eight categories on the bases of Biber et al. (1999) and Wang (2017), including NP fragment, PP fragment, AP fragment as modifier, AP fragment as complement, Clausal fragment, VP fragment, AdP fragment, and ConjP fragment. In congruence with Biber et al. (2004), both CCL and CEL lectures exhibit a mixed feature of 'oral' and 'literate' characteristics. An examination of non-standard FSs used by CCL lecturers indicates no communicative disturbance or breakdown arising in EMI classroom teaching.

Then, the FSs were functionally examined in connection with their corresponding moves/steps, where the majority of FSs were found to be directly or indirectly connected with the communicative purposes of the moves/steps. In terms of actual occurrences, Chinese lecturers used significantly less move signaling FSs than the native English lecturers.

The findings on rhetorical move structure and formulaic sequences are pedagogically conducive to Chinese EMI lecturers, pre-service and in-service novice lecturers in particular.

School of Foreign Languages   Student's Signature _Lin Deng_

Academic Year 2019   Advisor's Signature _A. Wannaruk_

ACKNOWLEDGEMENTS

The PhD experience has been a long journey full of thistles and thorns. I owe a special debt to my academic advisor, Assoc. Prof. Dr. Anchalee Wannaruk, who has always been around to escort me through the mist whenever I feel lost and depressed. Without her unfailing support and patient instruction, I wouldn't have survived the academic pilgrimage.

I remember the day when I first met her. She was such an elegant lady, hearing my introduction to the history of our city and enduring my nagging of immature research proposal for the program; I remember the day when I was desperately dismayed by her earnest disapproval of my research topic. She was such a thoughtful friend, sharing the hardship of her academic career and treating me with amazingly tasty Thai food; I remember the day when I was anxiously preparing for the proposal defense. She was such a true mentor of academics and life, commenting on the defense rehearsal over and over until perfection and taking all the trouble bringing me to the philosophical exhibit for inner peace. There are many more moments to share but for the limited space of the page.

She never stops giving me her continuous guidance, support and encouragement. She allows me immense freedom in academic thinking, which greatly boosts my confidence in the area of my academic pursuit. Her influence is present everywhere in this thesis, from theoretical architecture to technical details.

I would also like to extend my heartfelt gratitude to the committee members, Prof. Dr. Andrew Lian, Chair of the Committee, and three external examiners, Asst. Prof.

VII

I am greatly indebted to my beloved family. I could never have completed this thesis without their love and support. My mum Suhua Liu has always been wholeheartedly helping take care of my daughter and my newly-born son. My wife Na Li has sacrificed most of her time and energy shouldering my duty as a son, a father and a husband. I could never thank her enough. My daughter Lingxi and my little boy An'ze have always been the inspiration of my life. They constantly remind me of the meaning of life.

Finally, I would like to dedicate this dissertation to my late father Likang Deng, who always encouraged me in the pursuit of truth and knowledge. This is for you, Dad!

Lin Deng

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF FIGURES

LIST OF ABBREVIATIONS

AL	=	Additional language
APEC	=	Asia-Pacific Economic Cooperation
ASEAN	=	Association of South-East Asian Nations
BASE	=	British Academic Spoken English corpus
CA	=	Conversation Analysis
CARS	=	Create a Research Space
CBI	=	Content-based instruction
CCL	=	Corpus of Chinese Lecturers
CEL	=	Corpus of English Lecturers
CLCA	=	Corpus Linguistics and Conversation Analysis
CLIL	=	Content and Language Integrated Learning
DA	=	Discourse Analysis
EFL	=	English as a Foreign Language
ELF	=	English as a Lingua Franca
ELFA	=	Corpus of English as a Lingua Franca in Academic Settings
EMI	=	English-medium instruction
ESL/L2	=	English as a Second Language
ESP	=	English for Specific Purposes
EU	=	European Union
FL	=	Foreign language

LIST OF ABBREVIATIONS (Cont.)

IA	=	Interaction Analysis
IRF	=	Initiation-Response-Feedback
L1	=	First language
L2CD	=	Second language classroom discourse
LERN	=	Literacy and Education Research Network
LOB	=	Lancaster-Oslo-Bergen corpus
MICASE	=	Michigan Corpus of Academic Spoken English
MOE	=	Ministry of Education
NOW	=	News on the Web Corpus
NR	=	New Rhetoric
SFL	=	Systemic Functional Linguistics
T2K-SWAL Corpus	=	TOEFL 2000 Spoken and Written Academic Language Corpus
VBDU	=	Vocabulary-Based Discourse Unit
VOICE	=	Vienna-Oxford International Corpus of English
ZPD	=	Zone of Proximal Development

CHAPTER 1

INTRODUCTION

This introductory chapter presents an overview of the proposed study to compare the rhetorical move structure and formulaic language in university English-medium instruction (EMI) lectures given by native and non-native English lecturers. The chapter begins with a depiction of the research background, which sets the scene for the statement of the problem. Then it elaborates the rationale. After that, it states the research objectives and the significance of the study. The following sections of the chapter discuss the scope of the study, the relevant key terms, and the outline of the study. The final section ends the first chapter with a summary.

1.1 Research background

With accelerating internationalization and marketization of higher education in the past decades, universities are forced to compete on the global arena for more educational resources and larger market share. Primarily triggered by the aim of Bologna Process to standardize university degrees across Europe (Kirkpatrick, 2014), universities in Europe are offering more and more EMI courses and programs as a strategy to gain competitive edges in scientific innovation, institutional revenue, academic prestige and overall rankings. In a similar vein, the institutional aims of

regional organizations to facilitate regional cooperation, such as those of Association of South-East Asian Nations (ASEAN) and Asia-Pacific Economic Cooperation (APEC), greatly propelled EMI courses and programs in education sectors in Asia-Pacific countries, particularly at the tertiary level. EMI is gaining an unprecedented momentum in most parts of the world, in non-Anglophone Europe and Asia in particular. Recent statistics reports approximately 8,000 courses being offered in the English instruction medium at universities in non-Anglophone countries around the world (Mitchell, 2016, cited in Walkinshaw, Fenton-Smith & Humphreys, 2017). According to Crystal (2004), English has become the normal instruction medium in tertiary education in many countries, including those where English does not have the official status.

Despite widespread prevalence of EMI courses and programs, the concept of EMI remains an ill-defined and evolving term not fully agreed upon by researchers over the years (Airey, 2016). Instead of a 'monolithic fallacy' that assumes only one type of EMI, the British Council (2013) contends that EMI practices are highly context-dependent. In terms of socio-cultural contexts, EMI is implemented to teach non-language academic subjects in non-Anglophone "countries or jurisdictions where the first language (L1) of the majority of the population is not English" (Dearden, 2014, p.4). However, confining EMI to only non-Anglophone countries is problematic in the practical sense. On the one hand, EMI characterizes a large majority of, if not all, universities in Anglophone countries such as Australia where nearly all learning

activities are conducted through English (Walkinshaw, Fenton-Smith & Humphreys, 2017). On the other hand, higher education institutions in Australia and other Anglophone countries still face a wide range of EMI-related issues, e.g., limited exposure to the English language by international students even in the native English speaking environment, less than optimal English language proficiency outcomes of international students through the degree program (Humphreys, 2017). In other words, studying in Anglophone countries does not automatically lead to significant language proficiency improvement of the English as a second language (ESL) students, hence EMI in native English countries also merit attention. In addition, the term EMI in the literal sense may refer to any cultural contexts as long as the medium of instruction is English (Read, 2015). Therefore, EMI should encompass both Anglophone and non-Anglophone educational contexts (Walkinshaw, Fenton-Smith & Humphreys, 2017).

EMI in Chinese higher education is relatively new, and it only officially promulgated its debut at the turn of the new century in response to one of the Ministry of Education's (MOE) policy initiatives for improving the undergraduate program quality in Chinese universities. It requires that 5–10% of its undergraduate courses be offered in English or other foreign languages (MOE, 2001). EMI was recommended as a breakthrough strategy to "develop a global perspective in the Chinese university students, enhance their command of English, and provide access to cutting-edge knowledge in the West" (Hu, Li & Lei, 2014, p. 29). Furthermore, the Chinese MOE (2004) also included the number of EMI courses as a key performance indicator in

national assessments of higher education institutions. Aiming at further improving undergraduate education, several other ministerial directives and guidelines (e.g., MOE, 2005, 2007) also urge universities to offer more and better EMI courses and programs, and encourage overseas-trained Chinese academics and expatriate faculty to offer EMI courses at Chinese universities. In addition, many universities also provide benefits and incentives of various forms, such as favorable workload calculation formulae, subsidies, and institutional recognition, to encourage the faculties to undertake EMI courses (Tong & Shi, 2012). These policies at both national and institutional levels have greatly contributed to the upsurge of EMI in Chinese higher education.

1.2 Statement of the problem

Despite an exponential growth of EMI courses and programs in universities across China, a range of EMI-related problems in the Chinese context remain to be clarified and resolved.

Generally, EMI in China is a new initiative in education when compared with the situation in European countries. Although much empirical research on EMI has been conducted in Europe, little has been done in Asian countries, especially in China (Hu & McKay, 2012). Internally, research on EMI has mainly focused on theoretical and policy discussions, whereas empirical studies on real EMI classroom happenings have been generally lacking. Surprisingly little is known about the practices in EMI classrooms in the Chinese context, particularly the effects of EMI on students' subject

learning and language development. What makes things more complex is evidenced in the mixed and contradictory results regarding various aspects of EMI in studies conducted in European countries, e.g., faculty/students' supportive attitudes (Ball & Lindsay, 2013; Costa & Coleman, 2013, etc.) and resistance to EMI (Doiz, Lasagabaster, & Sierra, 2013; Tange, 2012, etc.), faculty/students' adequate (Unterberger, 2012, etc.) and inadequate English proficiency for EMI (Ball & Lindsay, 2013; Cots, 2013; Doiz, Lasagabaster, & Sierra, 2013, Wilkinson, 2013, etc.), no effect (Aguilar & Rodríguez, 2012, Hu & Duan, 2018, etc.) and detrimental effect on discipline learning (Hellekjær, 2010, etc.), as well as educational inequality issues involved in EMI practices (Costa & Coleman, 2013; Cots, 2013; Hu, 2009; Piller & Cho, 2013; Wilkinson, 2013, etc.). The findings are often confusing and sometimes misleading for Chinese educational practitioners.

In addition, some studies (e.g., Hu, Li & Lei, 2014; Hu & Lei, 2014) reported language difficulties encountered by EMI teachers and students. There appears to be an unbridged gap between the suboptimal language practices as were determined by students' and, to a lesser extent, teachers' less than ideal communicative competence in the English language, and the envisioned disciplinary and language learning goals of EMI programs. Despite various language support mechanisms by their university, both EMI lecturers and students still reported having difficulty in communicating the disciplinary knowledge effectively in English. They found it "challenging to use English to explain scientific concepts and technical terms, discuss the fundamental

processes and principles of their discipline, analyze complex cases, construct compelling arguments, and critique opposing views" (Hu, Li & Lei, 2014, p. 34).

1.3 Rationale of the study

The primary motivation to study university EMI lectures is closely related to suboptimal language practices in Chinese university EMI classrooms that are often attributed to the inadequate English proficiency of stakeholders in those situations. Language practices refer to the observable, regular, and predictable language behaviors and choices of both teachers and students (Spolsky, 2004) that constitute the EMI context in the real sense. Both teachers and students in EMI programs seem to lack the English proficiency for effective implementation of the EMI programs. However, it may be oversimplifying to attribute the difficulties in EMI classes to general English language proficiency. The focal university in Hu, Li and Lei (2014) and Hu and Lei (2014) adopted a range of language management guidelines for the EMI program, including recruitment and support mechanisms instituted for students and teachers, which in the actual sense guarantee high enough English proficiency among the teachers and students for effective implementation of EMI. In fact, most universities that offer EMI courses and programs adopted similar supporting measures for the sake of the courses and programs, so that they could ensure the quality of international programs, attract more international students, and improve their institutional rankings at home and abroad in the long run.

It is contradictory observing the self-claimed and other-rated language difficulties of students and teachers on the one hand, and their high English language proficiency on the other. Despite high English proficiency, students still encounter difficulties in comprehending academic lectures (e.g., Lebauer, 1984; Olsen & Huckin, 1990; Thompson, 1994; Allison & Tauroza, 1995; Flowerdew & Miller, 1997), and teachers find it challenging to deliver academic lectures through the medium of English (e.g., Hu, Li & Lei, 2014; Hu & Lei, 2014). It stands to reason that the difficulties in EMI lectures may not be a result of their general English proficiency, but instead something that is related to the instruction process of the discipline course through the English medium. Lecturers are equipped with all the necessary vocabulary and grammar knowledge, but they still find it challenging to give lectures in English. On the other hand, students may understand every single word of the lecturer, but they may still feel at a loss what the lecturer's main points are.

ESP researchers and teachers generally agree that the problem lies mainly at the discourse level, rather than the sentence level (e.g., Chaudron & Richards, 1986; Olsen & Huckin, 1990). Students in academic lectures are always faced with the daunting task of processing lecturers' lengthy talks in real time, combining new and old message, distinguishing between crucial and non-crucial information, when attempting to construct a coherent interpretation of the spoken data. Olsen and Huckin (1990) attributed non-native engineering students' difficulty in academic lecture comprehension to their failure in grasping the logical structure of lecturers' argument.

Meanwhile, it is perhaps no exaggeration that some lecturers are just not good at lecturing (Barr, 1990). Despite professional training in their own disciplinary fields, some lecturers receive little or no pedagogical guidance concerning how to organize lessons. Lecturers must simultaneously monitor words they have just uttered and prepare for the subsequent utterances appropriate for the circumstance (Brown & Yule, 1983a). Though lectures are typically well planned and prepared beforehand as a combination of memorization, aloud reading, and fresh talk (Goffman, 1981), they tend to show a dominance of the quasi-spontaneous fresh talk which involves a certain amount of real-time information processing and delivering during lecturing. Lecturers are always expected to make structured utterances that help students to build coherent mental representations of what lecturers are trying to say (Brown & Yule, 1983b).

In summary, the real reasons that lead to the suboptimal language practices in EMI classrooms, to some extent, may be attributed to the fact that students and lecturers are either unaware of the genre knowledge, or perhaps non-native lecturers may use a different set of lecture genre from that preferred by native English lecturers.

Studies on EMI lectures in China generally follow the strand of theoretical and policy discussions (Zhu & Yu, 2010; Hu, Li & Lei, 2014), but little attention has been given to the empirical investigation of the structural organization, or the generic structure, of university EMI lectures. No study has ever touched upon the potential differences in terms of generic structure of EMI lectures between native Chinese and native English lecturers. The present study intends to fill in this gap so as to shed light

on the EMI pedagogy in the Chinese educational context.

As a type of powerful discourse analysis approach, genre analysis provides a systematic analysis of the repeated communicative purposes and their linguistic exponents in a given genre. On the basis of the revised Create a Research Space (CARS) model, Swales (1990) posits the two-level rhetorical move structure schemata with three moves and varying numbers of steps for research article introductions. The genre analysis framework proposed by Swales (1990) is represented by the recurrent generic features, or moves with a number of steps, of a genre, as well as the lexico-grammatical features that realize them.

Ever since its advent, genre analysis has been overwhelmingly focused on written data. Lectures, as a unique spoken genre with their own communicative purposes and characteristic content and style, have generally been given little attention. In spite of paucity of research on genre analysis of university lectures, some (although very few) researchers did conduct genre-based analyses of lecture introductions (e.g., Thompson, 1994; Lee, 2009), lecture closings (e.g., Cheng, 2012), and the whole university lectures (e.g., Young, 1994; Lee, 2011, 2016). These studies have forcefully demonstrated the power of genre analysis theories, particularly the Swalesian move analysis framework, in disentangling and dissecting classroom discourse, which therefore is adopted as the theoretical rationale for the present study.

Equally important to the construction of coherent classroom discourse in university EMI lectures is related to the facility for lecturers to use multi-word sequences in the

instruction process. Psychologically, multi-word sequences might be stored and retrieved as whole prefabricated units, thus lessening the mental processing burdens of speakers (Wray, 2002), especially non-native lecturers who might be less equipped to deliver lectures in a freewheeling manner.

These multi-word sequences are prevalent in language use. According to Biber, Johansson, Leech, Conrad and Finegan (1999, p. 995), approximately forty-five percent of words in conversation and about twenty-one percent of the words in academic prose occur in such multi-word sequences, which they term as 'lexical bundles'. If lexical collocations are included, the multi-word sequences in native adult language could reach as much as seventy percent (Altenberg, 1990). In addition, multi-word sequences (e.g., *We will/'ll talk about*, *Let me turn to*) are often endowed with discourse structuring functions which are essential for generic analyses of discourse, both written and spoken. They are basic to language processing, production and comprehension (Wood, 2002; Schmitt & Carter, 2004; Meunier, 2012).

Given the importance and prevalence of such multi-word sequences, they merit careful examination of their roles in discourse construction, particularly in university lectures where both lecturers and students are under the tremendous pressure of real-time language processing.

It is worth mentioning that the predominant trend in this research area has been the frequency-based approaches adopted by Biber and his colleagues. Researchers in this strand rely on computer programs to automatically identify 'lexical bundles', i.e., "the

most frequently recurring sequences of words" (Biber & Barbieri, 2007, p. 264), "regardless of their idiomaticity, and regardless of their structural status" (Biber et al., 1999, p.990). Nevertheless, such frequency-based approaches suffer from the innate flaws of disregarding the structural and semantic unity and multifunctionality of multi-word sequences, which are not uncommon occurrences. Worse still, they overlook the infrequent but perceptually salient multi-word sequences (e.g., *you're never going to believe this*, in Biber, 2009, p.276) since such multi-word sequences were often considered non-idiomatic in previous studies. Therefore, instead of a frequency-based approach, the present study adopts intuition-oriented mixed diagnostic criteria to identify formulaic sequences with perceptually cohesive meanings or functions not easily discernible from their componential words.

In brief, the present study was mainly motivated by the palpable language difficulties encountered by teachers and students in Chinese university EMI lectures. A closer look at the underlying causes suggests that the suboptimal language practices might be attributed to the lack and/or the lecturers' different conceptualization of genre knowledge, and perhaps their insufficiency of formulaic language, rather than the less than ideal English proficiency in general, which is clearly not the case at least in university EMI classrooms in China.

1.4 Research objectives

In light of the problems and rationale stated above, the proposed study intends to probe into the real EMI classrooms and explore both the macro-structure and the micro-linguistic features of university EMI lectures given by native English lecturers and native Chinese lecturers, so that a comprehensive landscape of university EMI lectures across different cultures could be depicted for the benefits of teachers and students alike. A contrastive analysis of rhetorical move structure and formulaic language in EMI lectures by native and non-native English lecturers is expected to provide pedagogical implications for lecturers and students in the international higher education context.

The present study goes beyond the traditional tripartite opening-body-closing division of the classrooms, since this type of structural division is generally vacuous, and practically any speech event can be structured in this way (van Lier, 1988). The study endeavors to work out the rhetorical move structure of university EMI lectures selected from Michigan Corpus of Academic Spoken English (MICASE), the British Academic Spoken English (BASE) corpus, and self-built corpus of EMI lectures at a Chinese university according to their communicative purposes. Genre analysis of rhetorical move structure expects to sensitize both students and teachers in terms of dynamic information flow and structural organization of university lectures.

Apart from that, the study also attempts to investigate how formulaic language may help realize the communicative purposes of moves/steps.

In brief, the study aims to: 1) identify and compare the rhetorical move structure

of university EMI lectures given by native and non-native English lecturers; 2) identify and compare formulaic sequences used in university EMI lectures given by native and non-native English lecturers; 3) situate formulaic sequences in the rhetorical move structure and expound how formulaic sequences may affect the communicative purposes of the corresponding moves/steps.

These objectives are accordingly translated in three research questions as follows:

1) What are the similarities and differences in rhetorical move structure of university EMI lectures given by native and non-native English lecturers?

2) What are the similarities and differences in the use of formulaic sequences in university EMI lectures given by native and non-native English lecturers?

3) What are the potential connections between formulaic sequences and rhetorical move structure?

1.5 Significance of the study

The present study intends to investigate the rhetorical move structure and the formulaic sequences in university EMI lectures given by native and non-native English lecturers. It is significant in a number of aspects in various areas of English Language Teaching (ELT) and applied linguistics.

Firstly, as stated earlier, studies on EMI lectures in China are mainly theoretical and policy discussions (Hu, Li & Lei, 2014), lacking empirical research on classroom instruction processes. Unlike previous studies, this research collects and analyzes

classroom discourse from real university EMI lectures. It expects to present a real linguistic and discoursal landscape of what transpires in university EMI lectures in China.

Secondly, different from previous studies on lecture move structure (e.g., Young, 1994; Dudley-Evans, 1994; Thompson, 1994; Lee, 2009) that were motivated by the needs to improve students' lecture comprehension, this research is more concerned with difficulties of non-native English lecturers in delivering disciplinary content. Though not intended as a straitjacket of a genre model, the rhetorical move structure identified from MICASE and BASE lectures, as well as the Chinese EMI classrooms undoubtedly will have significant pedagogical implications for EMI lecturers.

Thirdly, a contrastive investigation of the rhetorical move structure of MICASE and BASE lectures and Chinese EMI lectures will reveal systematic differences embodied in instructional preferences that until now remain largely unknown, or relatively equivocal. A contrastive study of the rhetorical move structure is pedagogically invaluable, especially considering the perceived rarity of successful EMI experience in Chinese universities. Practically, Chinese EMI lecturers, pre-service and in-service novice lecturers in particular, are sure to benefit from such comparisons.

Fourthly, the present study is expected to produce a list of formulaic sequences, which might be, to a great extent, able to lessen the processing efforts for both teachers and students. The knowledge of formulaic sequences promises to help non-native English lecturers to deliver lessons in a more spontaneous and freewheeling manner.

On the part of the students, knowing more formulaic sequences expects to reduce their psychological burden in academic lecture listening and comprehension.

Finally, it is definitely worth mentioning the academic value of situating formulaic sequences in rhetorical move structure. Though some researchers have attempted to link formulaic sequences (lexical bundles/clusters) to their rhetorical moves in research articles (RAs) (Shi, 2014), RA introductions (Cortes, 2013), and RA discussion sections (Le & Harrington, 2015), they all focused on written data. In addition, these studies adopted a bottom-up methodology, starting from the identification of lexical bundles/clusters and analyzing them in their contexts to establish the relationship between the two categories (e.g., Cortes, 2013; Le & Harrington, 2015). That invites the flaw of disregarding those text segments where no lexical bundles/clusters appear, thus weakening the validity of the results to some extent. The intricate relationship between formulaic sequences and the rhetorical moves calls for more systematic elaboration and further clarification, and should be expanded to spoken data. This study will go to greater detail and depth in the hope of disentangling and demystifying the correlation, if any, between formulaic language and the communicative purposes of rhetorical moves.

1.6 Scope of the study

The present study attempts to make a contrastive analysis of both the rhetorical move structure and the formulaic sequences in university EMI lectures given by native

and non-native English lecturers. Therefore, the scope of the study is confined to the following areas.

1) Lectures from MICASE and BASE are selected to represent university EMI lectures given by native English speakers. MICASE is an academic spoken language corpus, covering a wide range of speech events in contemporary university settings. The BASE corpus consists of lectures and seminars collected from a variety of university departments, distributing across four broad disciplinary groups. Only 'science-oriented', 'small lectures' that are 'mostly monologic' by 'native speakers' are selected for the present study.

2) Lectures given by non-native English lecturers are collected at a polytechnic university in the northwestern province of Shaanxi, China. Nevertheless, it should be noted that EMI lectures in Chinese universities are often delivered in the medium of a code meshing of both English and Chinese (Bolton & Botha, 2015; Wu, et al., 2010; Zhao & Dixon, 2017; Hu & Duan, 2018). The present study focuses on EMI lectures that use only English as the medium of instruction. Therefore, EMI lectures given by non-native lecturers (in this case native Chinese lecturers) are confined to lectures given to international students where only English is used in classroom instruction.

3) In light of the complex nature of formulaic language, the intuition-oriented mixed diagnostic criteria were adopted in the present study to identify formulaic sequences, the details of which are discussed in Section 3.4.1.

1.7 Key terms

For the purpose of the present study, the working definitions of related key terms are given as follows:

1) EMI refers to the "use of the English language to teach academic subjects (other than English itself)" (Macaro et al., 2018, p. 37) in both Anglophone and non-Anglophone contexts (Walkinshaw, Fenton-Smith & Humphreys, 2017). In this research, EMI lectures include lectures given by native English lecturers selected from MICASE and the BASE corpus, and those given by Chinese lecturers from the self-built EMI lecture corpus.

2) Rhetorical move structure refers to the Swalesian segmentation (i.e., *moves* and *steps*) of text as defined by their communicative purposes.

3) A **move** "refers to a section of a text that performs a specific communicative function" (Biber, Connor & Upton, 2007, p. 23). Due to different transcription conventions in the spoken corpora involved, the present study didn't use sentences as the minimum unit for genre analysis. Instead, any text segments, be they complete sentences or sentence fragments, were considered moves so long as they convey distinctive communicative purposes. In cases where a text segment contains multi communicative purposes, only one move is identified in accordance with the most prominent communicative purpose.

4) A step is a lower level of text segment that provides more detailed linguistic means and options of realizing the communicative purpose of a move.

5) A formulaic sequence refers to a sequence of words "with a cohesive meaning or function that is not easily discernible by decoding the individual words alone" (Martinez & Schmitt, 2012, p. 304). A formulaic sequence does not have to be continuous so long as the sequence expresses a cohesive meaning or performs a certain holistic function. In this study, 'formulaicity', 'formulaic language' and 'formulaic sequence' are used interchangeably to refer to the same phenomenon.

6) Corpus of English Lecturers (CEL) is a corpus of 12 lectures selected from MICASE and the BASE corpus on the criteria of 'science-oriented', 'mostly monologic' (with the average 100-450 of words per turn (WPT)) (Simpson, Lee & Leicher, 2007, p. 34), 'small lecture' (with less than 40 students) (p.18), and 'native speaker'.

7) Corpus of Chinese Lecturers (CCL) is a self-built corpus of 12 EMI lectures given by Chinese lecturers to international students at a polytechnic university in Northwestern China. They are a collection of 'science-oriented', 'mostly monologic' (with average WPT 100-450), 'small lectures' (with less than 40 students) given to international students by Chinese EMI lecturers.

1.8 Outline of the study

The present study consists of seven chapters. Chapter one sets the background for the study. Chapter two presents a detailed literature review. Chapter three justifies the methodology for the research. Chapter four analyzes and compares the rhetorical move structure identified in CCL and CEL. Chapter five examines formulaic sequences in the

two corpora. Chapter six investigates the connections between formulaic sequences and their corresponding moves/steps. Chapter seven concludes with a summary of the findings, the implications as well as suggestions for further research.

1.9 Summary

This chapter begins with the introduction to the research background and research problems. Noticing an unbridged gap between language practices and the envisioned goals of EMI courses and programs in China, the present study intends to probe into real EMI classrooms and present an empirical investigation of the real happenings in EMI classrooms. Then, the rationale and objectives of the study are illustrated. Specifically, the present study adopts the Swalesian genre analysis framework to investigate the rhetorical move structure of EMI lectures given by native and non-native English lecturers. Meanwhile, formulaic sequences in each rhetorical move/step are analyzed in the hope of uncovering the potential connections between the two different categories. After that, the significance, the scope of study, the key terms and the structural organization of the study are clearly stated. In the next chapter, a review of the theoretical background and previous studies are presented.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter reviews previous studies and theoretical framework pertinent to the present study. It begins with a review of major analytic approaches to classroom discourse, with the purpose of illuminating their academic contributions and limitations, and at the same time setting the research at a broader theoretical background. Following that, genre theories are provided as the main theoretical framework for the present study. Then, the concept of English as a lingua franca is introduced as a new paradigm to investigate the practices involved in EMI lectures. On the basis of the literature reviewed, the next section briefs the proposed study and ends the chapter with a summary of literature review.

2.1 Classroom discourse

Classroom discourse refers to the language that teachers and students use to communicate with each other in classrooms. It is through classroom discourse that most teaching and learning takes place in classroom settings. Essentially, students "access new knowledge, acquire and develop new skills, identify problems of understanding, deal with 'breakdowns' in communication, establish and maintain relationships and so

on" (Walsh, 2011, p.2) through classroom discourse.

2.1.1 Language orientations of classroom discourse

Though excessively complex, heterogeneous and particularly messy in itself, classroom discourse remains the institutional interaction centering upon institutional goals. Seedhouse (2010, p.1) defines the institutional core goal of English as a Second Language (ESL/L2) classroom interaction as "the teacher will teach the learners the L2". And this goal remains constant whatever teaching method instructors might be adopting, whatever the students' language background might be and wherever the teaching is taking place. In ESL classrooms, "Language is both the vehicle and object of instruction" (p.2). In a similar vein, the core goal of content class interaction could be analogously expressed as *the teacher will teach the students the content knowledge through the use of language*. Obviously, language in a content class mainly serves as the vehicle or the medium of instruction. With the huge differences in mind, it is necessary to make a distinction between language classes, content classes as well as language-content classes when investigating classroom discourse.

Language classrooms mainly refer to ESL or English as a Foreign Language (EFL) classrooms where the language, in this case English, functions as both vehicle and object of instruction as indicated by Seedhouse (2004, 2010).

Content classrooms broadly refer to educational settings where the pedagogical concern is focused on the content, e.g., science, arts, and history lessons in primary and secondary schools, as well as lectures of all kinds in universities. Although

content classes mainly involve native speakers of English at different educational levels, recent years have witnessed more and more attention being given to ESL/EFL students who voluntarily, or are required to attend EMI subject classes of university level programs.

Despite the disputes and controversies of the term as discussed in Section 1.1, it is generally accepted that EMI is something that stand on its own and has no exclusively stated language-learning aims (British Council, 2013). Language gains are mostly considered as a by-product of the attainment of content knowledge in the subject course (Taguchi, 2014). Although EMI can be implemented at all educational phases, there is more EMI practice at tertiary level than at secondary level and primary level (Dearden, 2014). As a major channel of knowledge dissemination in higher education, university lectures provide a perfect venue for EMI practice, which aims at transmitting academic content knowledge through the language of English.

Closely related to the concept of EMI is another pedagogical approach that deals with both language and content concurrently, viz., Content and Language Integrated Learning (CLIL). Launched in Europe in the 1990s, CLIL emerged as a double-edged educational approach where an additional language is used for the instruction and learning of disciplinary content and language concurrently (Coyle, Hood & Marsh, 2010). Language is not the designated topic in CLIL classroom instructions, but there are language-related goals alongside the content-related objectives (Dalton-Puffer, 2007). Though an incarnation of the multilingualism policy

of European Union (EU), CLIL has been launched and fueled from both the high-level policy and grass-roots actions (Dalton-Puffer, Nikula & Smit, 2010) in mainstream education at primary, secondary as well as tertiary levels, which was probably allured by the envisioned double foci on both content and language.

CLIL is an attempt by European education specialists to establish their own unique framework to promote second/foreign language education, which differentiates itself from concepts and terms in North American bilingual education practices and content-based instruction (CBI). Unlike CBI, classroom instruction content in CLIL is not based on everyday life or general content of the target language culture, instead it is usually selected from disciplinary subjects or professions (Wolff, 2007). CLIL is also different from most other bilingual education in the sense that classroom instruction is usually conducted not by language experts, be they native or non-native, but instead by content experts to students who have already mastered basic literacy skills in their mother language.

Besides, CLIL also differs from the prevalent EMI practices. While EMI does not stipulate a specific contextual origin, CLIL is uniquely situated in Europe to promote plurilingual competence from among all EU citizens. A second difference lies in the linguistic concerns of the two pedagogical approaches. Different from the common practice of EMI, CLIL does not specify with which L2, foreign language (FL), or additional language (AL) the disciplinary content is to be integrated. Most importantly, though both CLIL and EMI take account of content and language in their

original envision and both are content-driven in practice, CLIL has a clearly-stated objective of cultivating both content and language as its own defining merit, whereas EMI does not have the unequivocal target.

In spite of all the differences, EMI and CLIL at least share a common ground of using English as the medium of instruction.

2.1.2 Approaches to classroom discourse analysis

The study of classroom discourse has long been the subject of exploration for researchers in applied linguistics, education, ethnography and ethnomethodology. The earliest systematic study of classroom discourse could be traced back to 1910 when researchers used stenographers to record the teacher-student talk in high school classrooms. With the application of audiotape recorders in classroom discourse research in 1930s, education specialists began to show immense interest in the impact of classroom discourse on learning (Hinkel, 2006). Since then, research on classroom discourse has been conducted from a range of diversified perspectives and approaches, among which four are particularly relevant to the present research, including discourse analysis (DA), conversation analysis (CA), combined approaches of corpus linguistics and conversation analysis (CLCA), and variable approaches.

2.1.2.1 Discourse analysis

Researchers in DA are interested in how language units beyond sentence level (in writing) or utterance level (in speaking) are situated in and affected by broader contexts. Therefore, discourse analysis of classroom interactions focuses on

words and utterances beyond the sentence level with the aim of examining their functions in contexts.

Following a structural-functional route to linguistic analysis, Sinclair and Coulthard (1975) developed a descriptive model of classroom discourse hierarchy: lesson → transaction → exchange → move → act. As the smallest discourse unit, each speech act may carry with it a specific function such as *request*, *explanation*, or *elicitation*. Then, these acts may work jointly to form the higher level of classroom discourse of move (similar to a 'turn' in conversation analysis), until the ultimate discourse unit of a whole lesson. While analyzing classroom discourse, Sinclair and Coulthard observed that classroom interaction between teachers and students displays an underlying initiation-response-feedback (hereinafter referred to as IRF) structure that remain the same in all classroom settings across different grades. Essentially, most classroom discourse follows such IRF three-move structure, which demonstrates how instructors manage and control the class.

However, the complexity of multi-party classroom settings renders it extremely difficult and even problematic to allocate naturally occurring patterns of classroom discourse to functions since most classroom utterances are multi-functional and dynamic. Another aspect of Sinclair and Coulthard system that often invites criticism lies in the fact that their recording was conducted in the traditional primary school classrooms in the 1960s where a clear status and power relationship between teachers and students could easily be identified. On the contrary, contemporary

classrooms often favor far more equality and partnership between teachers and students. So it remains doubtful as to the applicability of the framework of Sinclair and Coulthard (1975) to adequately describe the structure of classroom discourse today.

2.1.2.2 Conversation analysis

CA is based on the premise that social context is dynamically created and shaped by participants' conversational interactions, e.g., turn-taking, floor holding, openings and closures, sequencing of acts. Conversers' contributions to the interaction are inextricably linked to each other. One speaker's contribution is dependent on previous and/or subsequent speaker's contributions to generate new interactional contexts. Therefore, conversational interactions are not only *context-shaped*, but *context-renewing* (Walsh, 2011).

Though CA has its origin in the study of ordinary conversational interactions, it has strong parallels to such institutional discourses as classroom interactions. Applying CA to the analysis of classroom discourse, researchers are able to gain insights into how teachers manage classroom actions and create opportunities for educational purposes. Taking an emic perspective, researchers do not attempt to fit data into any preconceived categories, but instead focus on conversational patterns as they emerge from natural dynamic settings.

However, CA approaches are not free from their limitations. CA has been frequently criticized for the innate 'selectivity' and 'ungeneralizability'. The selection of data in CA analysis is considered contrived and idealized just to illustrate

a context-specific point, with little endeavors to put them in the whole conversational context. A natural consequence of the data selectivity is the inability to generalize findings to other conversational contexts.

Despite the criticism, CA proponents argue that the aim of classroom discourse research is not to generalize, but instead to facilitate understanding.

2.1.2.3 Combined approaches of corpus linguistics and conversation analysis

Though criticized for ungeneralizability, CA approaches remain a powerful way of analyzing interactional features of classroom discourse such as turn-taking, turn management, turn preference, etc., thus promoting understanding of discourse in educational settings. CA approaches provide deeper insights into the interactional dynamics of classroom discourse, but the results might be context-specific. That is, a specific result from CA research may not frequently recur, and in some extreme cases, may only occur once.

On the other hand, corpus linguistics (CL) studies are concerned more with the repeated occurrences and distributional patterns of words and word clusters. They are able to show a higher frequency of certain linguistic forms, but they don't go beyond the concordance lines where they are extracted, which typically consist of one or two sentences. The CL studies are, to some extent, de-contextualized.

Although with different research foci, CA and CL may complement each other to provide a fuller linguistic profile of classroom interactions. As Walsh

(2011) has argued, CL and CA should be combined since they are not only compactible to each other, but the synergistic benefits would enable deeper and richer descriptions and analyses of discourse. The combined methodology may remind researchers of certain high frequency linguistic forms that are worth attention through CL analyses, and meanwhile would provide a wider interactional landscape beyond sentences, therefore, engendering richer description and deeper understanding of classroom interactions.

2.1.2.4 Variable approaches

A number of studies (e.g., van Lier, 1988; Jarvis & Robinson, 1997; Kumaravadivelu, 1999; Seedhouse, 2004) adopt variable approaches to investigate classroom discourse, treating language classroom as 'real' constantly shifting social contexts. Classroom interaction patterns need to vary to adjust to different pedagogical goals and social relationships of the moment in that context. The variable approaches treat classroom discourse as dynamic interactions mutually constructed by the participants. For example, van Lier (1988) attempted to relate classroom activity to language use, and identified four types of L2 classroom interaction, viz., 'less topic-orientation, less activity-orientation', 'more topic-orientation, less activity-orientation', 'more topic-orientation, more activity-orientation' and 'less topic-orientation, more activity-orientation'.

Studies in variable approaches reveal finer variations of the classroom discourse, and enable deeper understanding of 'language socialisation'

(Lantolf, 2000, p.156). However, there always lacks a set of generally agreed nomenclature for describing and interpreting the micro-contexts in classrooms, with van Lier using 'types of interaction' (1988, p.156), Jarvis and Robinson six 'pedagogic functions' (1997, p.212), and Seedhouse four 'contexts' (2004, p.102).

2.1.3 Critique of approaches to classroom discourse analysis

Research on these approaches contributed to deeper insights into the complex communicative interactions in classroom settings, particularly in language classes. Specifically, DA focuses on classroom discourse above sentence level, and CA investigates patterns of classroom discourse in specific social contexts. Combined approaches integrate the strengths of corpus studies and conversation analysis, and variable approaches treat classroom as dynamically changing social contexts. There has been growing interest in the dynamic social dimension of classroom discourse.

However, it must be noted that all these approaches focus on language classrooms. Research interest in these approaches are closely linked to the institutional goal of these classes, i.e., language learning. Given the increasing internationalization of global education, research could go beyond just language education, and it should be extended to non-language education fields. Both CA and Variable approaches emphasize the pedagogical importance of classroom interactions on language learning, and they deem that classroom context is dynamically and mutually constructed by participants. However, in classes other than language, other dimensions, e.g., delivery of content, may outshine classroom interactions, though smooth and friendly

interactions may also help to achieve the institutional goals of these classes. In other words, the delivery of discipline contents is, at least, no less important than classroom interactions.

In addition, research in the DA approaches mainly examine the microcosm of IRF structures, which reveals little with regard to the positioning of discourse patterns in the larger classroom interaction cycle (Lee, 2011, p.19). Such research sheds little light on either how classes are organized at a higher discourse level and/or how participants in classroom interactions co-build the classroom discourse at various linguistic levels.

Lastly, the combined approach of CLCA integrates the strengths of Conversation Analysis and the power of Corpus Linguistics, which instantiates a promising combination of qualitative and quantitative methodology for the investigation of classroom discourse.

2.2 Genre analysis

The primary goal of this section is to establish Swalesean genre analysis as the theoretical framework for the investigation of classroom discourse.

2.2.1 Schools of genre theory

The last three decades or so has seen soaring interest in the concept of genre and genre-based pedagogy. Though originally a literary construct, genre has now developed into a popular framework for analyzing nonliterary discourse such as

research articles and an effective tool for course design and research in composition studies, professional writing, and English for specific purposes (Hyon, 1996).

However, genre has never been a clear-cut construct given the prevalent interest from a wide range of disciplines. Various research fields have explicated and explored genre in different ways. Although Flowerdew, J. (2002) insisted a linguistic vs. non-linguistic dichotomy in genre analysis research, most researchers (e.g., Johns, 2002; Hyland, 2004) have followed the three worlds of genre scholarship as distinguished by Hyon (1996). Hyon provided a holographic picture of three broad but overlapping approaches, viz., (a) Australian systemic functional linguistics (SFL)/Sydney School, (b) New Rhetoric (NR) studies, and (c) English for Specific Purpose (ESP), to genre scholarship with reference to their individual contexts, goals, and instructional frameworks for genre-based pedagogy.

2.2.1.1 Australian systemic functional linguistics/Sydney School

The Australian approach to genre studies, also known as the Sydney School in the United States, has its theoretical origin in systemic functional linguistics developed by Michael Halliday. Informed by studies of Malinowski (1935) and Firth (1957), Halliday (1978) argues that context of language use exerts great influence on meaning construction. Language users make meaning with reference to both the context of culture and the more immediate context of situation or 'environment of the text' (Halliday & Hasan, 1989, p.6). Fundamentally, SFL is concerned with "the relationship between language and its functions in social settings" (Hyon, 1996, p.696-697).

Different from formal linguistics that sees language as a set of rules of grammatical forms, SFL linguists view language as a social phenomenon of meaning making to communicate functions through linguistic choices in social contexts, allowing language users to express their real-world experiences, to interact with other language users, and to create coherent and cohesive messages (Hyland, 2004).

Halliday (1978) expounds that three essential metafunctions related to social context work together to bring meaning to text: the ideational, the interpersonal, and the textual. The ideational meta-function refers to the linguistic representation of human experience of the world. The interpersonal meta-function involves the interactive nature of language to build the relationship between language users. The textual meta-function helps organize a text into a piece of writing or speech in unified and coherent manners. These three meta-functions constitute the three broad dimensions of the central concepts of register in SFL, viz., field, tenor, and mode. Field refers to the activity that is going on, and it is related to ideational meaning. Tenor is the social relations of the participants in the activity, and it is related to interpersonal meaning. Mode refers to the role of language in an activity (i.e., the channel of communication), and it is related to textual meaning. To a great extent, any text produced, be it written or spoken, are dependent on and constrained by these dimensions of register of a language.

Hyland (2004) argues that texts are connected to contexts at both the level of register and genre. Building on Halliday's language theory, Martin, Christie

and Rothery (1987) developed theories of genre in the SFL framework, and described genre as a "staged, goal-oriented social process" with "structural forms that cultures use in certain contexts to achieve various purposes" (p. 59). It emphasizes "the purposeful, interactive, and sequential character of different genres and the ways that language is systematically linked to context" (Hyland, 2004, p. 25).

Genre is essential for effective and efficient communication since it provides a framework of what language users are expected to come across in a text. SFL view of genre has stressed not only the importance of social purposes of genre, but also of describing rhetorical move structure evolved to serve these purposes.

While register describes regularly occurring activities in a broad sense, such as *job applications*, *film reviews*, and *recipes*, in terms of their field, tenor and mode, genre is concerned with a set of more concrete conventions of rhetorical patterns for organizing texts, such as *narratives*, *recounts*, *arguments*, and *expositions*. A linguistic categorization of genre elements "allows analysts to see how these elements combine in different ways to make up the genres that are found in a range of different contexts and activities" (Hyland, 2004, p. 28).

Despite the preferences of SFL to tertiary education, genre-based pedagogy in Australia has been mainly concerned with primary school and secondary school L1 students, and more recently with adult migrant English education and workplace training programs. The SFL genre-based pedagogy aims to help "students become more successful readers and writers of academic and workplace texts" (Hyon,

1996, p. 700). It also takes as its goal to empower students, particularly those from minority and nonmainstream groups, with linguistic resources for social success, as well as to grant new migrants from limited educational and non-English-speaking backgrounds access to linguistic and social resources. In this regard, researchers have endeavored to link linguistic texts with functions in social contexts.

Drawing on the Vygotsky's (1978) concept of Zone of Proximal Development (ZPD), the area between what learners can do independently and what they can do with assistance, SFL has developed rich and sophisticated instructional frameworks implementing genre-based pedagogy. Particularly worth mentioning is the 'Teaching and Learning Cycle' promoted in the Literacy and Education Research Network (LERN) project in the late 1980s. This cycle depicts three phases in the process of genre instruction: modelling of text, joint negotiation of text, and independent construction of text. In the modelling phase, teachers present lexico-grammatical features, schematic structure, and functions of a text to students to help them model a variety of school genres, such as *narratives*, *procedures*, and *explanations*. The joint negotiation phase essentially involves a teacher-student negotiating process where teachers help record and shape students' contributions into texts that approximate the genre under discussion (LERN, 1990, p. 11). In independent construction stage, students are given opportunities to attempt to construct a model genre under discussion on their own. In addition, SFL researchers who are concerned with adult ESL instructions have also developed a genre-based competency framework for workplace

training programs, migrant English training programs in particular.

2.2.1.2 New Rhetoric studies

Greatly informed by postmodern social and literacy theories (e.g., Bakhtin, 1986) and North American research on first language (L1) rhetoric, composition writing, and professional writing (e.g., Freedman & Medway, 1994a), New Rhetoric (NR) studies embrace a number of perspectives that investigate genre mainly from social, cultural, and institutional contexts, and occasionally via the examination of the lexico-grammatical forms and rhetorical patterns. According to Miller (1984), genre is a form of social action that is "centered not on the substance or the form of the discourse but on the action it is used to accomplish" (p. 151).

Different from the stable and normative nature of language forms, generic forms in NR strands are "much more flexible, plastic, and free" (Bakhtin, 1986, p.79). Therefore, New Rhetoricians put greater emphasis on the dynamic nature of genres, which leads to "a far more provisional understanding of the concept" (Hyland, 2004, p. 35). In other words, genres are only 'stabilized-for-now' (though stabilized enough) (Schryer, 1993), open to change, and subject to negotiation. More broadly, NR researchers reject the possibility of transferring genre knowledge to classroom instructions since this involves seeking to "make solid what is actually shifting and variable" (Hyland, 2004, p. 39). In addition, classroom is viewed by NR researchers as an inauthentic context from the local socio-cultural settings. Given the dynamic, fluid, and blurred quality of genre, it would make little sense to extrapolate rules and

regularities from one context to another. In this regard, studying and teaching genres implies removing them from the contexts in which they make sense in terms of fulfilling social actions. Thereby, genres are, to some extent, mere artifacts for study rather than useful resources for effective communication (Hyland, 2004).

In line with their theoretical focus on socio-contextual aspects of genres, New Rhetoricians are not much concerned with the pedagogical application of genre theories for teaching linguistic structures of texts, but instead they have focused more on descriptions of situational contexts in which genres occur and the social actions/purposes that they fulfill within the contexts (Bazerman, 1988, 1994; Devitt, 1993; Freedman & Medway, 1994b; Miller, 1984, 1994). Explicit instructional frameworks for teaching students the linguistic features of genres have generally been lacking since NR researchers tend to view genres as guiding frameworks or rhetorical strategies rather than as recurring linguistic patterns. They are more interested in helping students and novice professionals understand the social actions genres accomplish as well as their corresponding contexts. (Hyon, 1996).

Bazerman (1988) suggests that the teaching of genre should not be concerned with the formal trappings of the genres, but it should aim at enhancing students' understanding of all of the 'life' embodied in texts. Freedman and Medway (1994c) echoes this preference of social contextual features to the linguistic and rhetorical move structure, arguing that "what has to be attended to . . . are features of the situation" (p. 11). NR scholars have focused mainly on describing genre features

and their contexts, and left with readers to decide and infer the pedagogical applications on their own (Hyon, 1996). According to Freedman and Medway (1994b), direct translations of genre knowledge into practical teaching are almost entirely absent. Instead of explicit genre instruction, NR scholars suggest that all teachers can hope to do is to endeavor to expose students to relevant genres, provide classroom activities and assignments that motivate students to respond in specified genres (Dias, 1994; Freedman, 1994; Hunt, 1994; Hyon, 1996), and inform them of overall features of format or organization (Freedman, 1994).

For NR scholars, the textual orientation of linguistic approaches (e.g., SFL, ESP) to genre over-emphasize the conventional form-function relations at discoursal level, and thus overlook potential creativity that may emerge within genres (Flowerdew & Wan, 2010, p.80). In light of that, they tend to employ ethnographic techniques to make 'thick descriptions' of contexts surrounding genres, e.g., observation, interviews, physical setting depictions, and text analyses (Hyland, 2004, p.37). As a consequence, NR scholars usually start with the social context and refer to linguistic regularities in texts to enhance comprehension of the context.

2.2.1.3 English for specific purposes

Researchers in ESP have been interested in genre as both an analytic and a pedagogical tool that can be utilized when teaching non-native English students in academic and professional settings. Drawing from an eclectic theoretical foundation, ESP embraces theoretical orientations from both SFL and NR (Belcher, 2006; Hyland,

2004; Johns, 2002). Though lacking "a systematic model of language" (Hyland, 2002, p. 115), the ESP approach to genre studies has been concerned with both text (spoken and written) structure properties and their communicative functions within social contexts. These concerns are explicitly reflected in the definition of genre by John Swales (1990), who asserts that "A genre comprises a class of communicative events, the members of which share some set of communicative purposes…In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience" (p.58). Essentially, communicative events are defined by both their formal structure properties and the communicative purposes as recognized and validated by expert members within social contexts. The academic works of Swales have greatly contributed to the understanding of these aspects of genre concepts. Specifically, Swales (1990) elaborated the linguistic rhetorical features of academic genre, and Swales (1998) detailed ethnographic descriptions of the social contexts in which genres are situated.

In analyzing texts, however, ESP scholars have been particularly interested in detailed descriptions of the formal characters of genres, but paid less heed to the social functions of texts as well as the social contexts (Hyon, 1996). These formal characteristics are recognized and validated by expert members of a discourse community as schematic structures due to their prototypicality revealed in a representative sample of texts. In turn, the schematic structures evolve into rhetorical moves and steps, the defining concepts in 'move analysis', as exemplified by Swales

(1990) in his seminal discussion of research article introductions.

Despite slight differences in the definition of 'moves' and 'steps' among ESP researchers (e.g., Bhatia, 1993; Nwogu, 1997; Yang & Allison, 2003; Kanoksilapatham, 2005), most would conceptualize 'move' as a distinctive communicative act to achieve a communicative purpose through a segment of text, while 'steps' usually act as a component of a 'move', providing more detailed linguistic means of realizing the rhetorical function of a move. Though a move may vary in length from a single sentence to several paragraphs, it generally contains one central theme, and "is designed to achieve one main communicative objective" (Swales & Feak, 2000, p. 35). In addition, "Both moves and steps may be optional, embedded in others, repeated, and have constraints on the sequence in which they generally occur" (Hyland, 2004, p. 47).

Eclectically pragmatic in its theoretical orientation, "the ESP approach is more linguistic than NR and more oriented to the role of social communities than SFL" (Hyland, 2004, p. 44). ESP researchers consider genres to be the property of discourse communities, and genre identification and analysis involve both the text-internal and text-external aspects, including socio-cognitive and social factors in particular communities rather than the culture at large. Since genres rarely exist in isolation, ESP researchers have also begun to relate genres with particular academic and professional settings. The concepts of 'genre sets' (Devitt, 1991, 1993) and 'systems of genre' (Bazerman, 1994) explicate how genres are networked to form a

'repertoire' of options for students in a particular academic or workplace context.

Apart from the exposition of rhetorical move structure and linguistic features of genres, as well as the ethnographic descriptions of academic and professional settings in which genres are identified, some ESP genre specialists (Swales, 1990; Weissberg & Buker, 1990; Bhatia, 1993; Flowerdew, J., 1993; Swales & Feak, 1994, 2000) also explicitly expound in detail the translation of genre to pedagogy. These pedagogies generally focus on rhetorical consciousness raising activities through classroom analyses of genres.

In summary, studies in the three schools of genre analysis have contributed to deeper understanding of the underlying connections between language forms and their communicative functions in social contexts. However, as stated above, they differ in their individual theoretical orientations, research methods, educational settings, and pedagogical applications.

The present study adopts the ESP approach as the theoretical framework to investigate the rhetorical move structure of university EMI lectures for the following reasons. Firstly, ESP embraces theoretical orientations from both SFL and NR, thus are eclectically pragmatic. Secondly, methodologically, ESP scholars are mainly concerned with text-internal analysis of rhetorical move structure and linguistic features, as well as text-external description of socio-cognitive and social factors in particular communities. In contrast, SFL mainly stresses the social purposes of genre, i.e., text functions in social settings, while NR generally adopts an ethnographic

description of social contexts. Thirdly, ESP research has focused on genre analysis in academic and professional training classrooms, therefore, are mainly concerned with non-native speakers at the university level. Nevertheless, SFL concerns mainly with primary school and secondary school L1 students and adult immigrants, whereas NR is committed to composition writing of L1 university students and novice professionals. Fourthly, while NR researchers reject the possibility of pedagogical application of genre theories for teaching linguistic structures of texts, ESP scholars have explicitly expounded in detail the translation of genre to pedagogy.

2.2.2 Contrastive rhetoric and genre analysis

Another strand of genre studies has been conducted from the perspective of contrastive rhetoric, a research area initiated by American applied linguist Robert Kaplan (1966), who originally examined differences and similarities in second language essay writing across cultures. However, the expansion of more genres in the purview of second language writing and acknowledgement of socially-situated nature of writing process have motivated scholars to adjust and supplement their research approaches to emphasize the social contexts of writing. As research methods for second language writing became increasingly context sensitive, contrastive rhetoric studies have evolved from mere text analysis to intercultural rhetoric approach that sees meaning construction as dynamic, socio-cognitive activities. Intercultural rhetoric researchers (e.g., Atkinson, 2004; Connor, 2004) called for a reconceptualization of culture to expand from the large 'received' static ethic culture in the national and/or

geographic sense to include small cultures such as classroom culture, disciplinary culture, and student culture, etc.

The development of genre analysis and reconceptualization of culture have expanded the realm of intercultural rhetoric to various academic and professional genres so that researchers can really "compare apples with apples" (Connor, 2004, P. 297). Contrastive rhetoric studies of genres have been generally conducted at the large national culture level (e.g., Zhu, 1997; Upton & Connor, 2001; Martín, 2003; Feng, 2008; Moritz, Meurer & Dellagnelo, 2008; Suárez & Moreno, 2008; Hirano, 2009; Loi, 2010; Loi & Evans, 2010; Li, 2010; Cho & Yoon, 2013; Lorés-Sanz, 2016) and the small disciplinary culture level (e.g., Morton, 2009; Lim, 2010; Basturkmen, 2012; Lin & Evans, 2012; Uhrig, 2012; Kuteeva, 2013; Stoller & Robinson, 2013; Kanoksilapatham, 2015; Tessuto, 2015; Kuteev & Negretti, 2016). These studies have revealed much about the cultural, both large and small, similarities and differences of various genre types, including sales letter, application letter, research articles, abstracts, etc. However, these are overwhelmingly confined to written data. Very little has been done on the contrastive analyses on spoken genres, university lectures in particular.

2.2.3 Corpus linguistics and genre analysis

Genre analyses of discourse are mainly concerned with identifying distinguishable discourse structure features with illustrative text excerpts, which may help navigate the grasp of the text as a whole, be it written or spoken. However, the results of these studies have been based on "comparatively small sets of textual data"

(Biber, Conrad & Reppen, 1998, as cited in Paltridge, 2012, p. 144), thus are not generalizable across contexts. On the contrary, corpus-based studies are capable of generating reliably generalizable results on the basis of large-size database analysis in a particular field of language use.

2.2.3.1 Trends in corpus linguistics

Currently, corpus studies "may look at language use in general, or they may look at the use of a particular linguistic feature in a particular domain" (Paltridge, 2012, p. 144), leading to two opposing trends in the compilation and application of corpora. On the one hand, corpora are getting astoundingly larger, with such general/reference corpora as the Bank of English, the British National Corpus, and the Cambridge International Corpus, expanding to the size of billions of words. Some on-line mega-corpora, like Corpus of Global Web-Based English, are expanding each minute. On the other hand, smaller, more specialized corpora, such as the Michigan Corpus of Academic Spoken English, the British Academic Spoken English corpus, the British Academic Written English (BAWE) corpus, and TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) Corpus, are being compiled to focus on specific genres of language use (Flowerdew, L., 2002; Koester, 2010).

2.2.3.2 Integration of corpus-based and genre-based approaches

Ever since the advent of the first corpora, corpus studies have long been criticized of the decontextualized nature of corpus data (e.g., Widdowson, 1998, 2002; Tribble, 2002; Hunston, 2002) as well as the atomized bottom-up analyses of data

(Swales, 2002).

However, advance in computer science has greatly facilitated the development of corpus linguistics. Corpus has now evolved into a powerful tool for discovering language patterns that otherwise may go unnoticed in the massive chaotic language data jungle. Flowerdew, L. (1998) contends that the exploitation of corpora has moved beyond the lexico-grammatical patterning of texts that aims at producing collocations and lists of fixed phrases, and is now more concerned with functional and rhetorical aspects of texts within various frameworks such as SFL, genre and discourse analysis. She advocates using specialized corpora to explore the classroom of academic/professional writing at a more discoursal level. Similarly, Paltridge (2001) and Wennerstrom (2003) also call for more integration of genre analysis and corpus-based investigations in discourse analysis of classroom language. In response to criticism against corpus linguistic analysis, Flowerdew, L. (2005) argues that corpus-based bottom-up methodologies and the top-down genre principles of text analysis could inform and benefit each other, and therefore should and could be integrated. In addition, McCarthy and O'Keeffe (2010) point out that applying corpus techniques in discourse analysis is not a one-way process since corpus studies can automate many (although not all) processes in discourse analysis while at the same time they also draw on discourse theories and applications.

In fact, some research, mostly conducted as early as the turn of the century, has already combined corpus-based and genre-based approaches in their

studies (e.g., Upton & Connor, 2001; Upton, 2002; Connor, Precht & Upton, 2002; Flowerdew & Dudley-Evans, 2002).

In response to Flowerdew, L.'s (1998) call for more research using corpus linguistic techniques to study text at discourse levels, Upton and Connor (2001) conducted corpus analysis to investigate the textlinguistic discourse moves of job application letters in the Indianapolis Business Learner Corpus (IBLC). On the basis of hand-tagged move analysis, the author utilized Wordsmith 2.0 to count and analyze politeness strategies used in Move 4 and Move 5 of the application letter genre through the linguistic features related to politeness. The authors argue that a textlinguistic approach considering the special features of specialized corpora is not only desirable, but also quite fruitful.

On the basis of studies of Bhatia (1997, 1998) on fundraising discourse, Upton (2002) categorized and tagged the rhetorical structure of a corpus of 242 direct mail letters from 71 organizations. Using the concordance program Wordsmith, the author analyzed and compared the frequency and lengths of each of the moves. This combined methodology perfectly incorporates the qualitative-oriented nature of genre analysis and the reliable quantitative power of computerized corpus analysis.

In another closely related study, Connor, Precht and Upton (2002) also demonstrate the value of combining genre analysis with corpus techniques. They manually tagged the rhetorical moves of 99 application letters of IBLC, meanwhile they

also used the concordance software program Wordsmith to automate the count of the occurrence of moves, which is almost impossible to do by hand considering the high frequency of these individual items in a corpus. They strongly advocate combining textlinguistic tools of genre analysis and corpus techniques so that broader statements could be made concerning rhetorical decisions writers make in a particular genre setting.

Flowerdew and Dudley-Evans (2002) conducted a genre analysis of 53 summative editorial letters to international contributors of the *English for Specific Purposes Journal*. On the basis of schematic structure examination, the authors also used corpus techniques such as frequency list and concordancing to extract the interpersonal dimension of communication. They also appeal for genre analysis of a corpus of editorial letters from a range of journals.

Other research also attempts to integrate these two approaches in the study of collocations in research articles (e.g., Gledhill, 2000; Marco, 2000), PhD literature review (e.g., Flowerdew & Forest, 2009), rhetorical structure of biochemistry research articles (e.g., Kanoksilapatham, 2005).

The combination of corpus linguistics and genre analysis produces richer level of description and generalizable results of various genre types. However, they are so far mostly confined to written data.

2.2.4 Previous studies on genre analysis of university lectures

A plethora of genre studies have been conducted to analyze rhetorical move structure and linguistic features in various academic and professional genres such as

research articles (e.g., Swales, 1981, 1990; Hopkins & Dudley-Evans, 1988; Nwogu, 1997; Holmes, 1997; Peacock, 2002; Yang & Allison, 2003; Kanoksilapatham, 2005, 2015; Lim, 2006, 2010, 2011; Parkinson, 2011; Amnuai & Wannaruk, 2013; Cotos, Huffman & Link, 2017), dissertations (e.g., Dudley-Evans, 1986; Hopkins & Dudley-Evans, 1988; Bunton, 2002; Kwan, 2006), abstracts (e.g., Salager-Meyer, 1990; Yakhontova, 2002), proposals (e.g., Connor & Mauranen, 1999; Connor, 2000; Halleck & Connor, 2006; Flowerdew, L., 2016), research reports (e.g., Nwogu, 1991; McKenna, 1997; Flowerdew & Wan, 2010) and business letters (e.g., Bhatia, 1993) etc.

However, perhaps due to a perceived messiness of spoken data, studies on spoken genre have been relatively few (Hyland, 2002), classroom discourse in particular. In light of the paramount significance in knowledge dissemination and the structured nature of classroom lessons (Richards & Lockhart, 1996), some (though quite few) researchers have conducted genre-based analyses on university lectures (Young, 1994; Deroey & Taverniers, 2011; Thompson, 1994; Lee, 2009, 2011, 2016; Cheng, 2012).

Since NR school mainly focuses on the ethnographic description of social contexts (e.g., Bazerman, 1988; Freedman & Medway, 1994c), little has been done on the description of linguistic features and rhetorical move structure of various genres in that strand. Therefore, the following section only reviews genre analysis of classroom discourse from the SFL and ESP perspectives.

Using the model of Systemic Functional Grammar, Young (1994) conducted

a comprehensive study on both the macro-structure and prominent micro-features of university lectures offered by native and non-native lecturers to EFL students across disciplines. A total of seven two-hour university lectures, with three given by non-native English lecturers from a Western European university and four by native English lecturers from a North American university, were collected and analyzed for the study. On the basis of line-to-line analyses of both semantic and syntactic choices of the classroom discourse, the author identified six phases (*discourse structuring*, *conclusion*, *evaluation*, *interaction*, *theory or content*, **and** *examples*) that form the macro-structure of lectures. Phases are strands of discontinuously recurring discourse throughout the lecture. Different from traditional division of a lecture into beginning/introduction, middle/body, and ending/conclusion, phases provide a more accurate configuration of the discourse structure of university lectures since there might be many beginnings, middles and ends in lectures as a result of recurrence of phases. The study focuses on the elaboration of the metadiscoursal phases (*discourse structuring*, *conclusion,* and *evaluation*) and their linguistic micro-features. In *discourse structuring* phase, a lecturer explicitly announces the direction of the lecture and tells students contents to be covered through the choice of particular verb groups (e.g., *give an example, give a list of words*, *give a description*, etc.), rhetorical questions, commands, as well as modals of prediction and intention. In *conclusion* phase, the lecturer makes a summary of points covered in the lecture through the predominant verbal group 'is', and the indicative declarative mood. Though less frequent than the

previous two phases, *evaluation* phase is also significant since in this phase the lecturer indicates explicitly to students his/her endorsement of and/or disagreement with the lecture content. This phase is often signaled by such judgment expressions as *very important*, *easiest and simplest way*, and *more direct way*, etc. The author draws a conclusion that there is a consistency of macro-structure across disciplines and between native and nonnative lecturers in university lectures, and pedagogically, descriptions of macro-structure and micro-features can benefit both teachers and students.

Deroey and Taverniers (2011) conducted a corpus-informed qualitative study of academic lectures. Though they didn't specify the theoretical orientation, they basically followed the SFL approach. According to the criteria of study levels, interactivity and audience size, the authors selected three lectures from each of the four broad disciplines of the BASE corpus to determine their communicative purposes. Guided by lexico-grammatical and contextual clues as well as existing lecture descriptions, they identified six main functions of academic lectures: *informing*, *elaborating*, *evaluating*, *organizing discourse*, *interacting* and *managing the class*, each comprising further sub-functions. The study also demonstrated the language patterns that help realize the discourse functions. Disciplinary variations were reported in the realization and prominence of informing and evaluating functions. Furthermore, some of the lecture discourse were frequently multifunctional.

In view of the considerable difficulty non-native speaker students experience in EAP lectures, Thompson (1994) used Swales' framework to analyze the rhetorical

features of lecture introductions in the hope of providing a robust generic model for more effective comprehension of lectures. Eighteen lecture introductions from a range of discipline areas were analyzed for their rhetorical functions according to the corresponding communicative purposes. Data analyses identified two main complementary rhetorical Functions, viz., ***Setting up the Lecture Framework*** and ***Putting Topic in Context***, each with several lower level Sub-Functions. In the identification of Functions and Sub-Functions, the study also examined their typical linguistic features. For example, the singular subject 'I' is a dominant feature indicating "the intrusion of the lecturer into the text" in ***Function Set up Lecture Framework***, whereas the personal pronouns 'inclusive we' and 'you' are used to show the lecturer's "recognition of the role of the audience's present knowledge and experience" in ***Function Putting Topic in Context*** (p.179). However, unlike the clear linear sequence reported in Dubois' (1980) study of biomedical conference speech introductions, no robust preferred orders have been found in the lecture introductions. One possible reason for the greater variation in rhetorical move structure of lecture introductions could be attributed to the pedagogical nature of lectures, where lecturers are under less pressure to exhibit control over conventionalized move structure to the academic inferiors of students. In addition, considering ephemeral nature of lectures, the author also cautions against forcing lecture introductions into "the straitjacket of a generic model" that leaves no room for flexible design and variable sequencing of lectures (p. 183). Finally, the author attempts to design some pedagogical activities to help students

predict lectures through the generic analysis of lecture introductions.

Following up on Thompson (1994), Lee (2009) examined the impact of class size on rhetorical move structure and linguistic features of lecture introductions. He compiled two corpora of lecture introductions of small-class and large-class lectures from MICASE. The study identified three moves and eleven strategies (steps) in the lecture introductions. While all moves occur in both types of lecture introductions, the strategies employed are different depending on the size of classes. Different from small-class lectures, the large number of students might have compelled experienced lecturers to use obligatory strategies of *housekeeping* and *looking ahead* to repeatedly remind students of class-related issues and upcoming lectures. In addition, while the strategy of *announcing the topic* occurs obligatorily in small class lectures, it occurs just once in almost all small lectures. However, large class lecturers may repeatedly resort to this strategy to keep the students in the loop. Meanwhile, the size of the class also seems to have effect on lecturers' use of personal pronouns, pronoun-integrated lexical phrases, and discourse markers.

Lee (2011) adopted a multidimensional, multi-perspective genre-oriented approach to explore the schematic organization and linguistic features of L2 classroom discourse, as well as its contextual dimensions. Through non-participant classroom observation and analysis of video-recordings of classroom interactions, the study found that, in spite of the spontaneity and improvised quality of classroom lessons, experienced L2 teachers may have internalized and formed stable language class

schemata that are recognizable by teachers and students. The study also used other data analysis procedures, e.g., semi-structured interview protocols and stimulated recall interviews, to triangulate the findings. Language classroom discourse was depicted as a distinct sub-genre of the classroom discourse genre proper.

Cheng (2012) investigated the rhetorical move structure, personal pronoun use, as well as the impact of class size in academic lecture closings. The study devised the framework of 'stages' and 'strategies' (similar to Swales' (1990) 'moves' and 'steps') to analyze the rhetorical move structure of lecture closings. Results indicated no preferred sequence of strategies across three stages in either large or small class lectures. Only a small proportion of small class lectures include all three stages, while over half of large class lectures contain only the *Ending Stage*. Strategies tend to cluster at the *Ending Stage* in large class lectures, but at the *Pre-ending* and *Post-ending Stages* in small class lectures. Personal pronouns were flexibly used by lecturers to encourage students' involvement, to establish rapport with students, and to mitigate disfavor in lectures. Examination of the correlation between personal pronouns and strategies revealed the impact of class size on the interactional dynamics of lectures.

Adopting a focused Swalesian move analysis combined with corpus-based methods, Lee (2016) made an exploratory study on the rhetorical move structure and linguistic features of classroom discourse of an EAP program designed for prematriculated ESL students. Meanwhile, stimulated recall interviews were conducted with the four EAP teachers in the program to generate an emic understanding of their

discursive practices. The study identified three major phases, viz., ***opening phase***, ***activity cycle phase***, and ***closing phase***, each with three moves and different numbers of steps. In addition, lexical phrases that signal discourse organization in each phase were also retrieved and analyzed. The ***opening phase*** mainly concerns with establishing positive classroom atmosphere and consolidating inter-class continuity. In this phase, lecturers frequently use 'we're going to/gonna' and 'I'm going to/gonna' to announce future lesson plans and actions. The ***activity cycle phase*** involves setting up and contextualizing activities for learning. Five lexical phrases (*you're going to/gonna*, *want you to*, *I want you*, *I want you to*, and *you want to/wanna*) occur frequently in this phase, and most of them are used for Move 4 Step 2: ***Outlining activity procedure***. In the ***closing phase***, a lecturer brings a lesson to an end and attempts to maintain good rapport with students. Three of the four lexical phrases (*want you to*, *you're going to/gonna*, *I want you*, *and I want you*) retrieved are overlapping, and they are mainly used for Move 7, Step 2: ***Outlining homework procedure***. The findings seem to suggest that EAP teachers have devised and internalized a stable lesson framework that may help organize their lessons in a more logical, navigable, accessible, and meaningful manner.

2.2.5 Previous studies on linguistic features of university lectures

The concept of genre generally consists of both the rhetorical/schematic structure and the language choices that are shaped by the communicative purposes of a genre. Although studies on the rhetorical move structure of university lectures have

been extremely few, research on linguistic features of university lectures abounds.

Research on classroom discourse in university lectures has been mainly focused on how various linguistic features are employed to help build positive learning environments to guide students in academic lectures. These features include personal pronouns (e.g., Rounds, 1987; Fortanet, 2004; Okamura, 2009; Yeo & Ting, 2014), metadiscourse (e.g., Thompson, 2003; Crawford Camiciottoli, 2004; Deroey & Taverniers, 2012; Lee & Subtirelu, 2015;), questions (e.g., Crawford Camiciottoli, 2008; Suviniitty, 2010; Chang, 2012), and formulaic sequences/lexical bundles (e.g., DeCarrico & Nattinger, 1988; Biber et.al, 2004; Nesi & Basturkmen, 2006, 2009; Biber & Barbieri, 2007).

2.2.5.1 Personal pronouns

Rounds (1987) remapped the semantic referents of personal pronouns of 'we', 'I' and 'you' in five English-medium university calculus classes given by native and non-native English graduate student teaching assistants. While examining the semantic mapping of the pronoun 'we', three more non-traditional semantic referents were identified apart from traditional 'inclusive-we' and 'exclusive-we', viz., the singular speaker 'I' (teacher), the addressee(s) 'you' (students), and generic 'we' (e.g., anyone in calculus or a specific field). It was found that teachers tend to avoid third person pronouns, but instead may turn to first and second person pronouns to include third person and indefinite reference. More than the concern of politeness, teachers may strategically use 'we' to help establish affective bonds with students in

the pursuit of ecology of consensuality in classroom interactions.

Fortanet (2004) investigated the frequency, linguistic contexts, referents and functions of the first person pronoun 'we' in English academic lectures. Contrary to the results of Rounds' (1987) study, 'we' is only used half as many times as other pronouns in the lecture and colloquium sub-corpus of MICASE. 'We' also displays a relatively low frequency in contrast to the first person singular pronoun 'I' in the mathematics talk sub-corpus of MICASE. That is considered to be a possible trend in spoken academic speech. Wordsmith Tools Concordancer analysis of the four-lecture sub-corpus of MICASE revealed that 'we' tend to co-occur with such verbs as 'have', 'talk', 'look', 'know', 'can/ be able to', 'need', 'going to', etc. On the basis of the linguistic context analyses, the author went on to analyze the referents and functions of 'we'. Nine references were identified, with the most occurrences of 'we' referring to a large group of people (including the reported speaker). Two discourse functions of 'we' identified include r*epresentation of groups* and *metadiscourse*. The study concludes that the various referents and discourse functions of 'we' could be discerned by the speakers' and hearers' negotiation of meaning through linguistic and extra linguistic clues.

Okamura (2009) examined the use of personal pronouns 'we', 'you' and 'I' in two speech events (9 undergraduate lectures and 9 public lectures) of MICASE. Results show that 'you' is the most commonly used personal pronoun in undergraduate lectures while 'I' is the most frequently used one in the public lectures.

Analysis of linguistic environments before the pronouns reveals that 'if' is used with 'you' much more frequently than other pronouns to indicate a hypothetical condition or engagement of audiences respectively in the two types of lectures. In terms of linguistic elements coming after the pronouns, 'you' tends to co-occur with verbs denoting students' activities in undergraduate lectures such as 'read', 'need', 'find', 'get', 'see' and 'look', while 'I' seems to go with verbs indicating speakers' opinions such as 'think', 'mean', 'guess', 'show' and 'hope' in public lectures.

Yeo and Ting (2014) investigated the use of personal pronouns 'we', 'I' and 'you' in 47 English-medium lecture introductions in the disciplines of arts and science in a Malaysian university. Results shows that 'you' is the most frequently used pronoun by the lecturers, followed sequentially by 'I' and 'we', which are similar in frequency. In terms of personal pronoun frequency, science lecture introductions exhibit higher density of personal pronoun use than arts lecture introductions. All three pronouns are found to be used for the two main functions of ***activating prior knowledge*** and ***giving instructions or announcements***. Functionally, science lecturers used more 'exclusive-we' and 'we for one' pronouns while arts lecturers tended to avoid the use of 'exclusive-we' and favored the use of 'you-generalized' pronouns when performing various lecture functions in lecture introductions. All these alluded to a disciplinary variation in the use of personal pronouns in lecture introductions. Science lecture introductions are more interactive whereas arts lecture introductions are more inclusive.

2.2.5.2 Metadiscourse

Metadiscourse, also known as language reflexivity, refers to discourse about discourse (Hyland, 2005; Ädel, 2010). Drawing on the work of Crismore and his collegues (1993), Hyland (2005) proposed the interpersonal model of metadiscourse, which is widely acknowledged and well-cited in various areas of discourse studies. Interpersonal model of metadiscourse covers both interactive and interactional resources. Interactive resources include transitions, frame markers, endophoric markers, evidentials and code glosses, and interactional resources consist of hedges, boosters, attitude markers, self-mentions and engagement markers.

Metadiscourse is common in everyday conversation as well as in specialized discourse, such as research articles, e.g., *In this paper, I explore…*, and academic lectures, e.g., *just to give you kind of a map of where we are going…* It has enormous potential to describe the way people organize ideas and the way writers and/or speakers relate to readers and/or listeners (Hyland, 2005). Metadiscoursal expressions help the writers/speakers explicitly signal the discourse structure and guide the audience through the discourse.

Thompson (2003) compared text-structuring metadiscourse and phonological paragraphs signaled by intonation between authentic lectures and EAP teaching materials. It was found, among the three different kinds of structuring metadiscourse markers (content markers, structuring markers and metastatements) at three different levels (global, topical and sub-topical), EAP material talks use more

metadisoursal signaling at the global level than the authentic lecturers do. Besides, there are also inconsistencies in terms of their respective use of metadiscoursal signaling among individual EAP material talks. On the other hand, despite shorter length, EAP material talks display 3.2 times more phonological paragraphs than authentic lectures do. Again, obvious differences were spotted between individual EAP talks. All these run the danger of implanting students with the unrealistic expectation of careful overall structure organization signaling from the lecturers, which clearly is not the case. The academic speakers use both metadiscourse and intonation signaling to help audience form a coherent 'mental map' of the overall structure organization of the talk. However, the EAP materials appear to be misleading considering the huge gaps both inside the EAP materials and between the EAP material talks and the authentic lectures.

Crawford Camiciottoli (2004) summarized three principal grammatical/lexical patterns of interactive discourse structuring: 1. PRONOUN + MODAL/SEMI-MODAL + MAIN VERB (e.g., *We will/'ll talk about*); 2. LET + PRONOUN + MAIN VERB (e.g., *Let me turn*); 3. PRONOUN + WANT + INFINITIVE (e.g., *I want to look*) (p.44). Then the author compared five self-collected guest lectures (given by two native lecturers and three non-native lecturers) in the University of Florence (L2 guest lecture corpus) with fourteen lectures selected from MICASE (L1 classroom lecture corpus) to investigate the role of interactive discourse structuring in lectures of different linguistic backgrounds and different levels of familiarity. The study didn't find interactive discourse structuring more frequent as

expected in L2 guest lectures. However, the highest frequency of discourse structuring among the non-native guest lecturers suggest that lecturers' linguistic background seem to exert more influence on the use of interactive discourse structuring than participants' unfamiliarity does.

Combining both quantitative and qualitative analytical procedures, Deroey and Taverniers (2012) presented a comprehensive overview of 'relevance markers' in 160 lectures from the BASE corpus. Relevance markers are metadiscursive devices that explicitly mark the relative importance or relevance of points in lectures. The study identified 782 instances of relevance markers, averaging about seven occurrences per 10,000 words. Depending on the main element, these relevance markers were structurally classified into adjective, noun, verb, and adverb patterns. Contrary to most intuitive understanding of the prototypical relevance markers, the most frequent ones are not the adjective patterns, but instead are the verb pattern V clause (e.g., *remember slavery had already been legally abolished*) and noun pattern MN v-link (a metalinguistic noun with the link verb *is*, e.g., *the point is*) devices that are often multifunctional. This discrepancy between stereotypical understanding and corpus linguistic reality of relevance markers sheds some light on EAP course design, teaching English for lecturing purposes, and related educational research.

Lee and Subtirelu's (2015) exploratory study compared the teachers' use of metadiscourse in second language classroom lessons and academic lectures. Based on the interpersonal model of metadiscourse (Hyland, 2005), the study searched

for and analyzed all potential metadiscoursal items from the self-collected L2 Classroom discourse corpus of 18 EAP lectures and the university academic lecture corpus of 18 MICASE lectures. Comparative analysis suggests that the context and content of instruction greatly influence the use of interactive metadiscourse in classroom teaching. While EAP teachers are concerned more with setting up various academic and linguistic tasks to engage students through explicit discourse frame markers, university lecturers lay more emphasis upon establishing relationships between disciplinary knowledge points. However, the use of some interactional metadiscourse features (e.g., hedges, boosters, attitude markers, self-mentions) didn't show much differences between the EAP classes and university lectures in the study. This seems to indicate a greater influence of the real-time spoken discourse nature of the classroom instruction than the disciplinary and pedagogical aspects in terms of interactional metadisourse use.

2.2.5.3 Questions

As the most common and direct form of human communication, questions have been a topic of interest to linguists from various research areas, e.g., the speech act theory, conversation analysis, systemic functional linguistics, etc. According to Thompson (1998), some questions are audience-oriented where the audience is at least symbolically offered an opportunity to reply, while others are content-oriented in which no audience response seems to be expected or really occurs. Questions have been investigated in various settings, including courtrooms (Levinson, 1992), press

conferences and news interviews (Clayman, 1993), radio phone-in programmes (Thornborrow, 2001), doctor–patient consultations (Ibrahim, 2001) and classrooms (e.g., Crawford Camiciottoli, 2008; Suviniitty, 2010; Chang, 2012). In classrooms, lecturers can point to important elements, organize the lecture into relevant sections and raise lecture interactivity through the use of questions.

Crawford Camiciottoli (2008) investigated the variation of questions used in academic lectures and written text materials. Two business studies corpora were built from twelve lectures and three textbooks plus online texts from two different websites for the research. It was found that unlike earlier studies, the frequency of question use in both lectures and written text materials is essentially the same. However, there does exist marked variation between the different communication modes in terms of the question forms and functions. Lectures show less frequent use of 'wh-questions' (12.6 vs. 22.6 per 1000 words) but more frequent use of 'yes/no questions' (14.9 vs. 5.3 per 1000 words) than written text materials. Questions in lectures showed features of both audience-oriented functions (reflecting face-to-face interaction) and content-oriented functions (reflecting the pedagogical aims) whereas questions in written text materials typically showed the content-oriented functions by the frequent use of 'wh-questions'. Additionally, it should not be neglected that significant variation was also identified among individual lectures, indicating lecturers' stylistic preferences on the use of questions. Internal comparison of the use of questions between the textbooks and online texts revealed an increasing degree of spoken/written hybridity on the latter,

highlighting the potential for the new innovative form of instructional media. In sum, this study shows that variation in the use of questions between lectures and text materials seems to "correspond more to the unique and evolving needs of instructional settings" (p. 1228).

Suviniitty (2010) examined the effect of lecturers' questions on students' perception of lecture comprehension in the ELF context of an English medium Master's program at a Finnish University. Based on 212 paper-based feedback questionnaires from students, the study compared three well comprehended lectures (numbered as L01-L03) and three less comprehended lectures (numbered as L04-L06) as rated in the student questionnaires. The major difference identified between the two groups of lectures, especially between L03 and L06, lies mainly in the use of questions. Though L03 and L06 were perceived as of nearly identical English level, students found L06, where hardly any questions were asked, more difficult to comprehend. In contrast, many questions, particularly genuine questions, were asked by the lecturer in L03, and it was rendered easier to understand. It seems to indicate that the use of questions tends to improve comprehension of lectures.

Chang (2012) examined the disciplinary impact on the professors' use of questions in academic lectures from a subset data of MICASE corpus. Despite slight individual differences, professors across the three disciplines seem to favor three question forms (out of the six): 'Wh-questions', 'Yes/no questions' and 'Declarative/Imperative + word tag questions', constituting over 95% of all the

questions in classroom interactions. No straightforward one-to-one mappings was found of the question forms and functions. Except for the sole-function 'tag questions', all other question forms exhibit a one-to-many and many-to-one interrelation with the question functions. Specifically, one question form may perform multi-functions, whereas a single question function can also be expressed by different question forms. The results seem to deny disciplinary variations of questions used by professors. Instead, the influence of university lectures as an established genre may exert more influence on classroom question use than disciplinary culture does.

2.2.6 Critique of previous studies on university lectures

In general, macro rhetorical move structure knowledge and mirco linguistic features as revealed in genre analysis of academic lectures are theoretically propelling for linguistic experts and pedagogically illuminating for educational practitioners. Most of these studies, except Young (1994) and Deroey and Taverniers (2011), adopted the ESP approach to investigate different lecture sections (lecture introductions: Thompson, 1994; Lee, 2009; lecture closings: Cheng, 2012) and whole lectures (Lee, 2011, 2016). These studies extend the scope of ESP research from written texts to spoken data, which further demonstrate the vitality and power of genre analysis framework. Comparatively, genre studies have paid much more heed to written texts than spoken discourse perhaps due to the complexity involved in recording, transcribing and analyzing the spoken data. However, the availability of more spoken corpora (e.g., MICASE, BASE, EFLA) has ignited more research interest in genre analysis of academic lectures. Researchers are

interested in not only the analysis of their own self-built corpora as in early studies (e.g., Young, 1994; Thompson, 1994), but also in the investigation of established ones (e.g., Lee, 2009, 2016; Cheng, 2012; Deroey & Taverniers, 2011). These studies have proven valuable in describing lectures as a genre proper, thus enriching the applicability of genre theories.

Pedagogically, all except Deroey and Taverniers (2011) mentioned the pedagogical value of genre analysis of lecture discourse. They all recognized the benefits of generic structure knowledge and linguistic patterns for teachers and students, especially pre-service and in-service novice teachers. However, only Thompson (1994) designed a series of concrete activities to guide students to predict lecture contents from lecture introductions. Perhaps due to different research foci, these studies didn't spare much space to expound the pedagogical applications of genre studies of academic lectures. Genre is not natural knowledge that all lecturers gain as soon as they step into the teaching profession, but something to be learned through professional training or perhaps accumulated through years of teaching in the profession in given educational contexts. Therefore, acquainting lecturers and students with genre knowledge of lectures seems necessary and urgent.

Furthermore, all except Young (1994) are confined to university lectures solely from native English countries. No genre analysis has been conducted on university EMI lectures in Chinese universities, and none has actually attempted to examine the possible differences of lecture genres in higher education sectors of

different cultures. In view of that, EMI lectures in Chinese universities deserve careful investigation.

Studies on the linguistic description of academic lectures have also greatly deepened our understanding of the complexities and dynamics of university lectures. Rich descriptions of the pronouns, metadiscourse and questions in lectures shed much light on unveiling the discoursive strategies experienced lecturers tend to use, which could be of great value to pre-service, in-service novice teachers, and students.

Nevertheless, these linguistic features can be of more practical value only when they are examined with their corresponding communicative purposes, i.e., the generic rationale. What's more, these linguistic features are of different categories, and are frequently overlapping. While pronouns are clearly grammatical terms, questions may be grammatical or pragmatic concepts. Metadiscourse is considered discourse structuring strategies in communication, and formulaicity can be a language processing strategy on par with language creativity (Wray & Perkins, 2000). A more practical way to study these linguistic features might be to investigate pronouns, questions, and metadiscourse embedded in formulaic sequences since they are extremely common in spoken discourse (Altenberg, 1998; Biber et al., 1999; Erman & Warren, 2000).

2.3 Formulaic sequences/lexical bundles

Most of the linguistic studies of lectures are, to some extent, intertwined with a growing research area that focuses on the use of multi-word expressions (hereinafter

referred to as MWEs, e.g., *in a nutshell*, *if you see what I mean*). Corpus studies have

demonstrated that MWEs are pervasive in natural language use (Altenberg, 1998; Biber

et al., 1999; Erman & Warren, 2000; Wray & Perkins, 2000; Biber et al., 2004; Schmitt

& Carter, 2004; Hyland, 2012; Meunier, 2012). Altenberg (1998), for example,

estimated that over 80 percent of words in London-Lund Corpus of Spoken English are

patterned in one form of such recurrent MWEs or another, which he referred to as

'recurrent word-combinations'. In another study, Biber, Johansson, Leech, Conrad, and

Finegan's (1999) analysis of Longman Spoken and Written English Corpus reported

approximately 45 percent of the words in conversation and 21 percent of the words in

academic prose occurring in expressions that they term as 'lexical bundle'. Not long

after that, Erman and Warren (2000) calculated that MWEs in the nomenclature of

'prefabs' accounted for 58.6 percent of the spoken English data and 52.3 percent of the

written data they analyzed. Despite varied statistical findings in these studies, all seem

to indicate the ubiquity of MWEs. On top of that, MWEs are considered building blocks

of coherent discourse (Hyland, 2008b) and important components for fluent language

output and successful language learning (Peters, 1983; Pawley & Syder, 1983; Wray &

Perkins, 2000; Mauranen, 2012).

However, in spite of general consensus on the significance of multi-word

expressions, surprisingly little was agreed upon with regard to their defining features,

identification methodologies, or even the technical terms to name them, hence different

operational identification criteria and varied research findings. At least 50 terms have

been used by researchers from different fields to refer to one or another facet of multi-word expressions, e.g., 'formulaic sequences', 'lexical bundles', 'lexical phrases', 'clusters', 'formulas', 'pre-fabricated patterns', 'prefabs' or 'chunks' (Wray, 2002). This reflects widespread interest in language formulaicity as well as its complex nature.

2.3.1 Profiling formulaic sequences

Multiple-word expressions have been defined and researched by researchers from different perspectives (e.g., Pawley & Syder, 1983; Nattinger, 1986; Werneit, 1995; Altenberg, 1998; Biber et al., 1999; Erman & Warren, 2000; Wray, 2002, 2008; Martinez & Schmitt, 2012), but most would agree MWEs are processed holistically the same way as individual words, serving certain functions and expressing specific meanings, though they usually consist of multiple orthographic or phonological words, fixed or semi-fixed to varying degrees.

A few points need to be emphasized to profile the concept of formulaic sequences. First of all, FSs are complex and multi-facet (Biber, 2009). It is extremely difficult to formulate a distinct FS category with clear-cut boundaries (Buerki, 2016, p.15). Generally speaking, formulaic sequences are treated as holistic MWEs, which may be due to their repeated occurrence (or high frequency), or the internal fixedness (or unpredictability), or sometimes both (Siyanova-Chanturia, 2013; Wang, 2018a; Wang, 2019). That is probably why FS has been suggested as an umbrella term to refer to a variety of multi-word expressions ranging from idioms (e.g., *beat about the bush*, *raining cats and dogs*), collocations (e.g., *take place*, *look up, French window*), clusters

or pre-fabricated expressions (e.g., *as can be seen*, *it has been noted that*), to phrase-frames (e.g., *the * of the study*, where the slot position * may be filled in by *purpose*, *goal*, *motivation* and *rationale*), which may vary greatly in terms of semantic transparency, morphosyntactic invariability, and structural completeness.

A natural consequence of the complex and multi-facet nature of FS is evidenced in the various identification orientations in the literature. Durrant & Mathews-Aydınlı (2011) summarized three main identification approaches to FSs.

1) The 'phraseological' approaches focus on the non-compositionality of fixed or semi-fixed expressions labeled as 'phrases' (e.g., *curry favor*, *French window*), 'idioms' (e.g., *spill the beans*, *kick the bucket*), 'prefabricated patterns' (e.g., *fully aware*, *highly significant*), as well as frame patterns with open slots to be filled (e.g., *too X to*, *as X as possible*). Due to semantic, lexico-grammatical and/or pragmatic concerns, these phraseological expressions are treated holistically, the meanings of which are not predictable from the constituent parts. The componential elements are not substitutable with words of similar meanings.

2) 'Frequency-based' approaches are more interested in the tendency for lexical bundles (e.g., *it is estimated that*, *for the most part*) to occur with higher than average frequency in texts of a register (e.g., Stubbs, 1995; Sinclair, 2004; Biber, 2009). Nevertheless, the frequency and dispersion threshold of lexical bundles are somewhat arbitrary. For example, Biber, Conrad and Cortes (2004) set the cutoff at 40 times per million words across five texts to identify 4-word lexical bundles, where Hyland (2008b)

set a minimum frequency of 20 times per million occurring in at least 10 percent of texts to extract 4-word lexical bundles. In addition, most lexical bundles are not idiomatic in meaning, nor are they complete in structure.

3) 'Psychological' approaches (e.g., Wray, 2002; Hoey, 2005) are concerned with efficient mental processing and storage of MWEs, which are dubbed under the term 'formulaic sequences' (e.g., *by and large*, *on the other hand*, *of course*). FSs are strings of words that language users remember and process as unanalyzable units rather than series of open slots for them to construct on-line with each use.

However, as suggested by Durrant & Mathews-Aydınlı (2011), the differences between these approaches should not be overstated, and they overlap to varying degrees. Non-compositionality and frequency-orientation may have led to the holistic processing and storage of FSs, whereas non-substitutability of FS constituents could have helped increase the FS frequencies.

Secondly, FSs develop the group identifying function, separating speakers and/or writers of a given discourse community from those that do not belong (Biber et al., 1999; Wray, 2002, 2008; Biber et al., 2004; Biber, 2006; Hyland, 2008b, 2012). FSs are

"retrieved whole from memory at the time of use" (Wray, 2002, p. 9) to meet communicative needs in a specific discourse community. The appropriate use of one FS form instead of another grant the language user the membership of that discourse community. Therefore, genres and registers across various discourse communities may,

to a certain extent, be characterized by FSs used in those communities (Biber, 2006; Biber et al., 1999; Kashila & Heng, 2014; Wang, 2017). Kashila and Heng (2014), for example, found that academic lectures rely heavily on FSs to fulfill functions related to their disciplines, with politics lecturers favoring topic elaboration and/or clarification bundles to make coherent links between ideas while Chemistry lecturers using slightly more stance and referential bundles. Wang (2017) also identified disciplinary preferences in the use of four-word lexical bundles across ELF lectures from different academic fields.

Thirdly, FSs are often employed by language users to express different meanings and realize various functions in discourse. For example, they can be used to signpost discourse organization, make evaluation, as well as transact routinized meanings (e.g., *Tell me about it!* expressing strong agreement). In fact, it has been suggested that there always exists conventionalized language (i.e., formulaic sequences) for every recurrent communicative need that arises from various genres (Nattinger & DeCarrico, 1992).

Lastly, FSs have long been found to have the processing advantage and are able to improve the overall language production. For example, it was found L1 readers (sometimes L2 readers) read FSs more quickly than the non-formulaic structures (Siyanova-Chanturia, Conklin & Schmitt, 2011; Underwood, Schmitt & Galpin, 2004; Conklin & Schmitt, 2008). Ellis and Sinclair (1996) noted that the attainment of language fluency involves the acquisition of memorized linguistic sequences for both

native and non-native language users.

2.3.2 Previous studies on formulaic sequences in university lectures

Recent development in corpus studies has revealed the pervasiveness of FSs in natural language use and their important role in academic discourse (e.g., Biber et al., 1999; Biber et al., 2004; Hyland, 2012). The use of FSs are reported to be able to help achieve fluency, facilitate comprehension and production, as well as identify membership (Wang, 2018b). However, much attention has been paid to one aspect of FSs or another in written discourse from various perspectives, including pedagogical orientation of FSs (e.g., Wray, 2000; Wray & Perkins, 2000; Simpson & Mendis, 2003; Durrant, 2009; Byrd & Coxhead, 2010; Simpson-Vlach & Ellis, 2010; Durrant & Mathews-Aydınlı, 2011; Crossley & Salsbury, 2011; Martinez & Schmitt, 2012; Cortes, 2013; Martinez, 2013; Staples, Egberta, Bibera & McClairb, 2013; Hsu, 2014; AlHassan & Wood, 2015; Morley, 2015; Peters & Pauwels, 2015; Allan, 2016; Cai, 2016; Bestgen, 2017; Hammond, 2018; Lu, Yoon & Kisselev, 2018), L1 vs. L2 variations (e.g., Durrant & Schmitt, 2009; Nekrasova, 2009; Chen & Baker, 2010; Ädel & Erman, 2012; Perez-Llantada, 2014; Qin, 2014; Salazar, 2014; Pan, Reppen & Biber, 2016; Grabowski & Juknevičienė, 2016; Lu & Deng; 2019), FSs in different research article sections (e.g., Le & Harrington, 2015; Lu, Yoon & Kisselev, 2018; Wright, 2019), disciplinary variations (e.g., Cortes, 2004; Hyland, 2008b; Durrant, 2017; Dong & Buckingham, 2018; Hyland & Jiang, 2018), spoken vs. written variations (e.g., Biber, Conrad & Cortes, 2004; Biber & Barbieri, 2007; Biber, 2009; Carey, 2013), novice vs.

expert writer variations (e.g., Hyland, 2008a; Ansarifar, Shahriari & Pishghadam, 2018; Wang, 2018a), and even ELF perspective of FSs (e.g., Kecskes, 2007; Mauranen, 2009; Carey, 2013; Martinez, 2018).

However, among the huge bulk of literature on formulaic sequences, relatively small number of studies (DeCarrico & Nattinger, 1988; Biber et.al, 2004; Nesi & Basturkmen, 2006, 2009; Biber & Barbieri, 2007, Neely & Cortes, 2009; Kashiha & Heng, 2014; Schnur, 2014; Carey, 2013; Wang, 2017, 2018b) have been concerned with FSs in academic lectures.

DeCarrico & Nattinger (1988) investigated lexical phrases in academic lectures on a variety of topics, covering anthropology, biology, ecology, etc. Lexical phrases are 'chunks' of functional language, which are used as higher level discourse signaling macro-organizers, e.g., *as it were* and *as X would have us believe*. The study identified two levels and eight functional categories of lexical phrases: a) global macro-organizers indicating overall structure of the lecture; b) local macro-organizers highlighting supporting information at specific points within the overall framework. Striking differences were found between 'Conversational Style' and 'Reading Style' lectures in all categories except global macro-organizer *Topic Markers*. The study concludes with suggestions on ways of teaching lexical phrases so as to foster students' lecture comprehension abilities.

Biber, Conrad and Cortes (2004) adopted a frequency-driven approach to examine lexical bundles in two important university registers: classroom teaching and

textbooks. They compared the lexical bundles in classroom teaching and textbooks to those identified in their previous research on lexical bundles in conversation and academic prose. It was found that classroom teaching reflects mixed characteristics of conversation and textbooks/academic prose. Nevertheless, instead of being an intermediate between conversation and academic prose as expected, classroom teaching uses the most lexical bundles in total types and tokens, as well as in each of the three functional categories identified, including stance expressions, discourse organizers, and referential expressions. This clearly reflects the complex communicative demands of the classroom context where instructors need to organize the information-oriented classroom discourse in an involved manner under the real-time production pressure.

Extending the research line of Biber, Conrad and Cortes (2004), Biber and Barbieri (2007) investigated lexical bundles across a wide range of university registers. Contrary to previous studies that indicate higher frequency of lexical bundles in spoken discourse than in written discourse (Biber, Conrad, & Cortes, 2004; Biber et al., 1999, chap. 13), this study found that lexical bundles are very common in written course management among all the university registers concerned. It suggests that lexical bundle use may be influenced not only by the communication mode, but also strongly by the communicative purposes in the register. In terms of classroom teaching, intention/prediction stance bundle are common only in this spoken register. They are often used to announce new topics and activities or predict future events. In addition to stance bundles, classroom teaching also relies heavily on discourse organizing bundles

and referential bundles.

Nesi and Basturkmen (2006, 2009) investigated lexical bundles in 160 university lectures of four different disciplines from BASE and MICASE. The study examined four-word lexical bundles that occur at least ten times within each individual discipline and over 50 times in the whole corpus to ensure wide dispersion and high frequency while avoiding idiosyncrasy. It was found that the majority of lexical bundles serve discourse signaling functions, which fall into two broad categories: 'referential expressions' that signal relationship between ideas and information in the lecture such as *apposition*, *contrast/concession* and *result/inference*, and 'discourse organizers' that flag relationship between topics and activities in the lecture, such as *topic introduction/ focus*. The study concludes with appeals to acknowledging the cohesive role of lexical bundles in language learning and teaching.

Neely and Cortes (2009) examined the functions of five topic-introducing/discourse organizing bundles (i.e., *if you look at*, *a little bit about*, *a little bit of*, *I want you to*, and *I would like you*) as identified by Biber et al. (2004) and Nesi and Basturkmen (2006) in academic lectures from MICASE. They found more functions of these lexical bundles than identified in previous studies. For example, out of the 54 times of occurrence of the bundle *if you look at* in the instructors' speech, only 17 instances (31%) was used as topic introduction, the other two-thirds being used for topic elaboration or clarification. The study further explored designing academic listening lesson plans to include explicit corpus-based lexical bundle teaching through

sample EAP classroom activities that raise students' awareness of discourse organizing bundles and guide them to discover the fundamental differences between these discourse organizing bundles in textbook (contrived) lectures and authentic ones.

Kashila and Heng (2014) investigated the disciplinary variation of four-word lexical bundles in eight university lectures from the BASE corpus (four Politics and four Chemistry lectures). They set the lexical bundle cutoff frequency at ten times per hundred thousand words occurring in at least three different lectures. More lexical bundles were found in Politics lectures (131tokens and 32 types) than Chemistry lectures (94 tokens and 26 types). Functionally, Politics lecturers rely more on topic elaboration or clarification bundles (e.g., *the way in which* and *come up with a*) to help make coherent links between ideas. On the contrary, Chemistry lecturers use more directive bundles (e.g., *you have got to*) to clarify the instruction or task, and more referential bundles to refer to a variety of abstract and concrete entities (e.g., *on the basis of*, *in the form of*), and materials and/or instruments (e.g., *one of the things*, *is a kind of*, and *a particular kind of*), which essentially involve experiment-oriented disciplines such as Chemistry. The study concluded that there were marked disciplinary variations between the Politics and Chemistry lectures in terms of the lexical bundle functions, suggesting lecturers' heavy reliance on bundles to fulfill discipline-related functions.

Schnur (2014) examined four-word discourse organizing lexical bundles in EAP lectures and authentic lectures. Four EAP listening textbooks covering six

disciplines were collected and compiled into an EAP lecture corpus with 104 simulated EAP lectures. Fifty-six lectures from Biber et al.'s (2004) T2K-SWAL teaching sub-corpus were used as samples of naturally occurring authentic lectures. Seventeen discourse organizing bundles identified from EAP lectures were not included in Biber et al.'s (2004) bundle list, whereas twenty-one of the discourse organizing bundles of Biber et al. (2004) were found to be infrequent in the EAP lectures of the study. Functionally, all three bundle function categories (i.e., topic introduction/focus, topic elaboration/clarification, and macro-level discourse organization) were significantly more common in EAP lectures than in authentic lectures, which corroborates previous studies that claimed greater reliance on discourse organizing bundles in recorded EAP lectures due to material writers' preference for explicit discourse organizing language in listening textbooks.

Carey (2013) investigated the frequency effects on the entrenchment of formulaic chunks and the distribution patterns of high-frequency formulaic organizing chunks in academic spoken and written ELF contexts. It was found interactive organizers *in my view* and *from my point of view* occurred relatively infrequent in both the spoken and written ELFA corpora and MICASE, which was deemed evidence of weaker entrenchment of the chunks that led to relatively high rates of approximated chunk forms (deployed in the same environments and with the same organizing functions, but with a fuzzier representation of the actual string of words), accounting for 40 percent and 19 percent of total forms in their respective corpus. The frequency

effect of *say* versus *speak* could have contributed to the preference of the approximate chunk *so to say* instead of the conventional form of *so to speak*. On the other hand, although spoken EFL discourse tend to drop articles, the high-frequency text organizing chunk *at the same time* primarily appears with the article in the conventional form in both spoken and written ELF corpora. Similarly, the highest frequency textual organizer *on the other hand* rarely occur in approximate forms. Taken together, these formulaic organizing chunks appear at much higher frequency in the ELFA corpus, indicating that ELF users are processing these chunks holistically as native speakers do. However, the author cautioned that the use of approximated chunks in the ELF contexts should not be overstated, as the majority of the chunks examined do conform to conventions, both in function and form.

Wang (2017) conducted a corpus-based study to investigate the effects of genre (i.e., academic lectures vs. seminars) and discipline (i.e., Medicine, Social Sciences, Natural Sciences) variations on the use of 4-word lexical bundles in the ELF context. Thirty-one lectures and thirty-two seminars from ELFA, amounting to a total of 508,769 running words, were used for the study. With reference to Biber et al.'s (1999, p. 994) findings on ENL speakers' bundle use, the author argued that ELF speakers' more frequent use of bundles and little deviation in terms of the lexical and syntactic formation may suggest their tendency to rely on the 'idiom principle' (Sinclair, 1991) in academic settings. Structurally, seminars strongly favor clausal bundles while lectures seemed to feature NP and PP bundles, suggesting lectures inclination to literate

discourse whereas seminars the opposite direction. Functionally, seminar discourse displayed greater proportion of participant-oriented bundle use whereas lecture discourse exhibited twice as many real-world oriented bundle use. In addition, it was also found that Natural Sciences lectures could be distinguished from Medicine and Social Sciences lectures in the use of collective '*we*' bundles as in '*we are interested in*' and '*we are trying to*'. On the other hand, both Natural Sciences and Medicine seminars featured bundles involving negative personal states such as '*I don't know X*' while Social Sciences seminars highlighted affirmative bundles involving '*I think*' to express personal opinions (e.g., *I think this is*). These results demonstrated that genre and discipline are two important factors in understanding academic ELF communication.

Using the lecture subset data in Wang (2017), Wang (2018b) established well-designed manual identification criteria and used UAM Corpus Tool to examine FSs that signal discourse organization in academic ELF lectures from disciplinary perspectives. It was found that the majority of the high frequency FSs are highly fixed two- or three-word sequences. Different FS use preferences were revealed between lectures from the three disciplines of Natural Sciences, Social Sciences and Medicine. Specifically, Natural Sciences sub-corpus showed more frequent use of discourse-structuring FSs than the other two discipline sub-corpora, particularly the spatio-temporal FSs (e.g., *first of all*, *with regard to*, *I'll show you*). The Social Sciences sub-corpus features use of FSs of elaboration (e.g., *which means (that)*, *I mean*, *let's say*, *that is*, *for example/instance*), logical connection (e.g., *that's why*, *because of*, *as long*

as), and transition (e.g., *on the other hand*). On top of that, the study also indicated that

variant forms of FSs identified (e.g., *as you see* vs. *you can immediately see*) were more

evident in semantically transparent and syntactically flexible sequences, indicating the

operation of the open-choice principle alongside the idiom principle in academic ELF

settings.

2.3.3 Critique of previous studies on formulaic sequences/lexical bundles

As clearly shown above, studies on formulaic sequences/lexical bundles in

spoken genre, in this case university lectures, are generally lacking. Though not always

explicitly stated, the MWEs generated in these studies are of great pedagogical value

for not only lecturers, pre-service and in-service novice lecturers in particular, but also

EFL/ESL students who have to attend lectures instructed through the medium of

English. Some studies actually attempted to explore the pedagogical application of

these MWEs. DeCarrico and Nattinger (1988) tried to integrate lexical phrase macro-

organizers in reading and vocabulary classes for ESL students, and Neely and Cortes

(2009) strived to design a series of lesson plans to instruct some of the topic-introducing

and topic-elaboration bundles to help ESL/SFL students with academic lecture listening

comprehension. Nonetheless, pedagogical aspects of formulaic sequences/lexical

bundles still call for more intensive research.

Another aspect of formulaic sequences/lexical bundles that deserve our

attention lies in the identification criteria. The majority of the literature reviewed

overwhelmingly adopted the frequency-based approach. They set various threshold

cutoff frequency and range to extract formulaic sequences/lexical bundles from the corpora. This has long been criticized for the arbitrariness involved in the extraction process, and consequent disunity in the comparison of research findings. Besides, many researchers doubt the value of the frequency-based criteria of MWEs since frequency itself is not an adequate guide to formulaicity (Wray, 2002). And this approach has always been attacked for its disregard of the structural and semantic unity and multifunctionality of lexical bundles, and overlooking of discontinuous and infrequent sequences that are pragmatically and pedagogically important (Wang, 2018a).

In terms of the research area, most have been concerned with MWEs across different genres/registers and disciplines. The results demonstrate much about lexical bundles used in lectures and other genres/registers, e.g., textbooks, conversation, and academic prose, as well as disciplinary variations of lexical bundles. In addition, some studies, though very few, have noticed lexical bundle use in the ELF context, e.g., Carey (2013), and Wang (2017, 2018b).

2.4 Connecting formulaic sequences and rhetorical moves

Though many studies have long been investigating linguistic features that may characterize or specify different rhetorical moves of a certain genre (e.g., Swales, 1981; Brett, 1994; Williams, 1999; Kanoksilapatham, 2003; Yang & Allison, 2003), empirical studies that link particular linguistic features to moves/steps have been limited (Cortes, 2013). Nevertheless, recent years have witnessed increasing interest in the connections

between multi-word expressions of various categories and rhetorical moves. Some have been interested in examining bundles in specific moves/steps (e.g., Durrant & Mathews-Aydınlı, 2011; Morley, 2014, Shi, 2014), while others are concerned with extending lexical bundles to the rhetorical moves/steps where they appear (e.g., Csomay, 2012; Cortes, 2013; Le & Harrington, 2015; Omidian, Shahriari & Siyanova-Chanturia, 2018), and still others focus on the genre-based pedagogy of lexical phrases (e.g., Cai, 2016; Mizumoto, Hamatani & Imao 2017).

2.4.1 Previous studies on bundle/formulaicity-move connections

Durrant and Mathews-Aydınlı were perhaps the first to systemically investigate formulaic language under Swales' (1990) framework of genre analysis. Recognizing the insufficiency of 'form-first' approaches, where frequency was taken as the guide to internal connections and communicative functions of phrases, Durrant and Mathews-Aydınlı (2011) adopted a 'function-first' approach, in which formulaic language in academic writing was identified and analyzed with reference to the communicative functions of the context. They adopted a top-down approach, and compared the use of formulaic language in 94 students' essays of British Academic Written English Corpus with that in 94 research articles from prominent journals. A multi-step iterative process was first executed to identify the moves and steps in the introduction sections of essays and articles in the two corpora. Then the indicating structure (IS) steps was specifically analyzed to demonstrate the extent to which this step was used by students and article writers, and how it was instantiated through

abstract constructions and lexical formulas forms. Though the IS step was more common in student essays than in articles, the use of this step as well as the linguistic forms to realize it show disciplinary variations in research articles, but not in student essays. In terms of the lexical formulas forms, research articles tend to be more formulaic in this step than student essays, and some parts of the lexical forms are highly formulaic whereas other parts are more diverse.

Drawing on the Swalesean genre analysis approach, Morley (2014) examined phrases used in 100 postgraduate dissertations completed at the University of Manchester. The researcher summarized numerous rhetorical patterns similar to Swales' 'moves', and used them as one of the main organizing sub-categories to develop the Academic Phrasebank. Similar to the IMRD major section division in most studies on research articles, Morley dissected students' dissertations into six main parts, including Introducing work, Referring to literature, Describing methods, Reporting results, Discussing findings, and Writing conclusions. Various rhetorical patterns, e.g., *Establishing the importance of the topic for the world or society*, were identified under each of these major dissertation sections. On the basis of the psychological insights into language learning and production, the researcher affirmed the phraseological nature of academic language and attempted to exemplify and highlight the phrases under each of the rhetorical patterns through a number of self-designed criteria. And the researcher made explicit the importance of instructing these phrases in the context.

Adopting the model of Kanoksilapatham (2005), Shi (2014) conducted a

contrastive study of moves and lexical bundles in English journal articles published in China and internationally. The author self-built two English research article corpora for the study, with forty-five RAs published in China and forty-five published internationally in the discipline of agricultural science in each corpus. The results indicated similar move structure between RAs in Chinese and international journals, with each identified sixteen moves. Nevertheless, some discrepancies were also spotted in the choices of move/step frequency between the two corpora, particularly in the Introduction and Discussion sections. In addition, the study also revealed distinct move patterns of agricultural science RAs, particularly in the Methods and Results sections, suggesting disciplinary variations of move structure in RAs. On the basis of that, the study identified lexical bundles in the sixteen move sub-corpora of each RA corpus. The results showed greater use of lexical bundles in the international corpus, indicating the necessity of explicit instruction of lexical bundles for Chinese researchers who wish to publish internationally.

Despite illuminating pedagogical implications in the top-down approaches to formulaic language research, the tremendous amounts of texts in large corpora renders it technically difficult to manually segment the discourse texts. Realizing the methodological limitation of the top-down perceptual evaluation on the segmentation of discourse texts in large corpora, Csomay (2012) and her research team devised a text processing tool to automatically divide classroom discourse into lexically coherent discourse units on the assumption that changes in linguistic patterns correspond to

variations in discourse functions. The author examined the distributional patterns of lexical bundle functions in small discourse units that they termed as 'Vocabulary-Based Discourse Units' (VBDUs) to explore their relationship with discourse structure. Results showed that 84 lexical bundles in Biber et al. (2004) occurred at least once for a total of 1,152 occurrences in the first six VBDUs. Specifically, in the opening phase of classroom discourse (the first three VBDUs), stance bundles appeared most prominently while referential bundles occurred the least frequently. The first three discourse units exhibited grammatical features as substantiated by lexical bundles that are closely related to class management such as details on take-home exams or future tests (e.g., *is going to be*, *I want you to*, *you don't have to*). VBDU 4, the start of 'instructional phase', which featured an informational focus associated with pedagogical and communicative purposes, revealed a dramatic drop of stance bundles, a spike of discourse organizer bundles signifying topic introduction and focus (21 per cent), and high percentage of referential bundles expressing attribute specification (26 per cent) and those indicating place, time, and text reference (11 per cent). VBDUs 5 and 6 displayed patterns of lexical bundle use similar to the VBDU 4. These findings indicated a strong correspondence between bundle functions and discourse structure through linguistic variation analysis (Csomay 2005).

On the other strand, Cortes (2013) investigated lexical bundles as well as their connections with moves in research article introductions. The author first identified 3849 tokens of 135 lexical bundle types in the corpus of research article

introductions from various academic disciplines. Then these bundles were analyzed structurally and functionally with reference to the taxonomy of Biber et al. (2003, 2004). Then she went on to analyze the lexical bundles in context to explore the communicative functions they convey or help to convey, i.e., specifying the moves/steps where the bundles appeared. She found that some bundles were exclusively linked with certain moves or steps (e.g., *As one of the*, *a great deal of*, *In a variety of* appeared only in ***Move 1 Step 1: Claiming relevance of field***), while others appeared in several moves and steps (e.g., *a wide range of* and *in the absence of* occur in three different moves and steps). In addition, longer lexical bundles of more than five words (e.g., *the purpose of the present study*, *it has been suggested that*) were found to be always performing as triggers that initiate particular moves/steps, while other shorter bundles (e.g., *with respect to the*, *the effect of the*) were also used as complements, accompanying the triggers and adding commentary to the main communicative function of a particular move or step.

Le and Harrington (2015) also explored the link between phraseology and rhetorical structure in the bottom-up approach. They built word clusters with fourteen keywords form the comparison of the Discussion section and the remaining sections of 124 quantitative research articles in applied linguistics. Moves and Steps were subsequently established on the extended context centering upon clusters. Then they investigated the distribution of these word clusters in individual steps of the ***Commenting on results*** Move of the Discussion section. They found that the most

common linguistic features used in the three steps of the ***Commenting on results*** Move were mainly hedging devices, including modal verbs (e.g., might, may), epistemic verbs (e.g., suggests) and modal adjective (e.g., possible). With regard to rhetorical move structure, the study revealed move cycles of the ***Reporting results*** Move co-occurring with each of the Steps of the ***Commenting on results*** Move, together with several instances of move embedding. The findings have immense pedagogical values and provide support for the link between phraseology and genre.

Omidian, Shahriari and Siyanova-Chanturia (2018) examined multi-word expressions in different rhetorical moves of research article abstracts across disciplines. A one-million-word corpus of 5910 abstracts from both soft sciences and hard sciences were built for the research. A frequency-based extraction of the corpus identified 198 4+-word lexical bundle types and 5759 tokens. The lexical bundles were examined quantitatively between soft science and hard science disciplines in terms of their general functions with reference to the taxonomy of Biber et al. (2004) and Hyland (2008b). Then each bundle token was analyzed in its immediate context to determine the rhetorical move it belongs to according to the framework by Dos Santos (1996) and Pho (2008). A qualitative analysis of the bundles in relation to the moves where they occur was subsequently conducted to examine disciplinary variations. The results revealed significantly more use of lexical bundle in terms of both types and tokens in Move 3 (***methods***) by hard science abstract writers, indicating their focus on promoting their research through statement of methodology and materials involved. On the other

hand, soft science abstract writers used significantly more lexical bundle tokens in Move 2 (***presenting the research and its purpose***) and Move 5 (***discussing the research and its findings***), suggesting their interest and effort in ensuring clear expression of the research goals and implications to the diversified target readership through the use of conventionally accepted bundles.

Cai (2016) focused on incorporating the genre-based pedagogy of Sydney school and the corpus-informed instruction into the ESP genre framework. The author conducted an exploratory study on the instruction of academic lexical phrases (LPs). A number of instruments and tools were adopted to measure students' genre knowledge and lexical phrase knowledge through pedagogical intervention. On the bases of corpus extraction and previous studies, the author selected 43 LPs for the intervention instruction of the study. Then, the teaching and learning cycle of Sydney School were adapted and implemented to help scaffold the learning of LPs. Finally, the author attempted to link LPs with genre in larger contexts according to the ESP genre framework. In addition, the author also proposed to differentiate move-specific lexical phrases (MLPs) from general lexical phrases (GLPs) since they serve different pedagogical purposes. The results reported significant growth among students in their receptive LP knowledge and genre structure, which indicated the effectiveness of the integrated approach for the instruction of LPs. More importantly, the study showed an interdependent connection between MLPs and target GLPs, and the correlation between genre knowledge and such linguistic features as lexical phrases.

Not long after that, Mizumoto, Hamatani, and Imao (2017) explored to develop a data-driven and theory-based online writing support tool for research articles. Adopting Pho's (2013) move framework, the researchers manually tagged the moves in all sections of the 1,000 research articles in their self-built corpus of applied linguistics, where they identified 25 moves in total. Then they performed correspondence analysis to empirically examine the correspondence between lexical bundles and moves, the results of which confirmed a robust association between the two constructs. Based on the effective proof-of-concept evidence, they devised a web-based writing support tool that combines moves and lexical bundles. Preliminary user feedback from L2 writers in Japan indicated beneficial effects the bundle–move connection approach aims to achieve.

2.4.2 Critique of previous studies on bundle/formulaicity-move connections

Researchers have been interested in the connections between moves and linguistic features since the very early days of genre theories (e.g., Swales, 1981). However, there has been a general lack of empirical studies to examine the connections between rhetorical moves/steps and lexical bundles/formulaic sequences/multi-word expressions. Durrant and Mathews-Aydınlı (2011) were among the pioneers to systematically investigate the move-phrase connections. They took a 'function-first' approach to identify the moves/steps of textual segments first, and then they attempted to specify abstract constructions and lexical formulas forms that were used to realize such move functions. Their work is in line with the function-first concepts of genre

analysis. However, they didn't make it clear how they identified the abstract constructions that were essential to dissecting such connections.

On the same strand of this top-down approach also include Morley (2014) and Shi (2014), who also started from the identification of moves and then went on with analyses of phrases and lexical bundles. It is undeniable such connections on the basis of the textual proximity may, to a very large extent, prove to be true. Nevertheless, it is not tenable to equate textual proximity to bundle-move connections and rush to claim the legitimacy of such connections.

On the bottom-up strand, Csomay (2012) resorted to computer programs to automatically identify the lexically-coherent macro-structure of classroom discourse. The research may have its algorithmical strength, but it is in conflict with the defining features of moves/steps as stipulated by Swales (1990, 2004). Cortes (2013), Le and Harrington (2015), and Omidian, Shahriari and Siyanova-Chanturia (2018) started their research from the identification and analysis of lexical bundles/phrases/multi-word expressions and then conducted genre analysis on the moves/steps established from their extended contexts. Methodologically and etymologically, studies on this strand inappropriately took textual proximity for the bundle-move connections. They mistakenly approve the bundle-move connections on the mere fact these moves/steps were established centering upon these bundles/phrases. On top of that, in her study of research article introductions, Cortes (2013) also identified a group of exclusive lexical bundles which she recognized as occur in only one step of each move. This was both

methodologically and pedagogically not reliable. The so-called exclusive bundles extracted from research article introductions may, to a great extent, appear in other moves/steps of other sections of research articles. For example, such bundles as *are likely to be*, *at the end of*, *as a result of the* were identified as exclusive to ***Step 3: reviewing items of previous literature*** in research article introductions. But they are very also likely to appear in the moves/steps of the discussion sections of research articles. This methodological flaw would perhaps jeopardize or undermine the pedagogical values of the research.

As for the genre-based pedagogy of lexical phrases, Cai (2016) focused on incorporating the Sydney school pedagogy and the corpus-informed instruction into the ESP framework. This theoretical integration strengthened the theoretical validity of the research which produced positive growth in lexical bundle use among foreign learners in their academic writing. However, they didn't even touch upon the question of the bundle-move connections. Instead, the author just presupposed such connections before they were empirically proven. Mizumoto, Hamatani and Imao (2017) are careful enough to use correspondence analysis to analyze the bundle-move connections before they develop on-line support tools for research articles. However, the correspondence analysis itself just shows the existence of the connections between the two variants, i.e., lexical bundles and moves, without specifying the strength degree of the connections. In addition, this statistical method needs to be complemented with direct analysis on the bundle-move connections *per se*.

Despite profound insights from previous studies, most research has mistaken textual proximity for bundle-moves connections. More empirical evidence is needed to establish such connections, which we know from our instinct may very likely exist and these connections could be strong in intensity. Empirical studies that directly address the move-bundle/formulaicity connections are urgently needed.

2.5 English as a lingua franca

The growing popularity of EMI has thrown English into linguistically-diverse contexts, where it is increasingly used as a lingua franca, i.e., a vehicular language for people who do not have a common native language in communication (Mauranen, 2003).

2.5.1 Conceptualizing English as a lingua franca

Although earlier studies have excluded native speakers from ELF (e.g., Firth, 1996; House, 1999), scholars today generally accept ELF as "any use of English among speakers of different first languages" (Seidlhofer, 2011, p.7), including native English speakers who may also use ELF as an additional linguistic resource for intercultural communication.

Instead of a standard English variety within the nation-state boundaries or a uniform and fixed mode of communication, ELF is a flexibly co-constructed and variable means of communication (Bowles & Cogo, 2015) in a 'community of practice', where participants are united by common practices of talking, beliefs and values in joint

activities (Eckert, 2000). The ELF users are not homogeneous as specified in the traditional sense of speech communities, but instead comprise people from different lingua-cultural backgrounds for common engagement of activities. Both the community of speakers and the location of the ELF encounters could be changing without being associated with any specific nation. This variability of ELF is locally constructed in specific events in different geographical locations and domains. Nevertheless, in the era of global communication, ELF encounters do not have to be confined to geographic proximity, instead they can be achieved in virtual communities via the use of social media applications (e.g., Twitter, Facebook, Wechat, QQ) on internet (Cogo, 2012; Jenkins, 2014; Mauranen, 2012, 2014). The intrinsic variability and contingent fluidity of ELF, therefore, establish itself as *sui generis* in its own right rather than failed attempts to approximate the native speaker norm (Jenkins, 2011).

University EMI lectures for international students in China make perfect settings for ELF communicative practices. The sole legitimate language for classroom instruction is English, the students and lecturers are of different lingua-cultural backgrounds, and the common enterprise of the lectures includes knowledge dissemination and acquisition, as well as facilitating academic, disciplinary and/or professional socializing of students (Deroey & Taverniers, 2011).

2.5.2 Previous studies on academic lectures in English as a lingua franca settings

To date， linguistic inquiry of ELF has been mainly conducted on the

following aspects: phonology (e.g., Jenkins, 2000, 2002, 2005), lexicogrammatical/ morphosyntactic features (e.g., Seidlhofer, 2004; Breiteneder, 2005, 2009; Cogo & Dewey, 2006; Erling & Bartlett, 2006; Hülmbauer, 2007, 2009; Dewey, 2007; Björkman, 2008a, 2008b), and pragmatics (e.g., Firth, 1996; House, 1999; Meierkord, 2000; Cogo & Dewey, 2006; Metsä-Ketelä, 2006; Ranta, 2006; Cogo, 2009; Klimpfinger, 2009; Mauranen, 2009; Björkman, 2010, 2011). Since the study intends to examine the rhetorical move structure and formulaic language of university EMI lectures, research on phonology, though essential for classroom instruction, is not the concern here, hence is excluded from the literature review.

Seidlhofer (2004, p.220) summarized eight categories of lexicogrammatical features of ELF, including dropping 3rd person -s, confusing relative pronouns who and which, Omitting definite and indefinite article, etc. These non-native-like language use or errors, which are considered in urgent need of correction in the English as a foreign language paradigm, appear to cause no disturbance in intercultural communication. Seidlhofer (2004) also revealed that with the use of such pragmatic strategies as topic change, rephrasing, repetition, or let-it-pass principle, violations of ENL pragmatic norms seldom lead to unintelligibility, and misunderstanding is infrequent in ELF interactions.

Ranta (2006) found the non-native like use of verb progressives such as '*are belonging to*' and '*we are breathing*' in the ELF context creates no confusion, but instead is creative usage of language resource to serve the 'attention-catching' function

to draw interlocutors' attention. Similarly, Mauranen (2009) also reported that the systematic use of non-standard expressions such as '*in/on my point of view*' and '*I'm not very sure*' causes no misunderstanding in ELF communication, thus only indicates the emergence of new innovative ELF patterns. On top of that, Metsä-Ketelä (2006) identified three new functions of the vague phrase '*more or less*', viz., minimising, comparing similarities, and approximating quantities. These new functions are also indications of innovative ways of using the language for new communicative functions.

These studies generally accept that non-standard language use rarely lead to confusion or communicative breakdowns, instead it is creative use of the language resource for communicative effectiveness. In addition, Cogo and Dewey (2006) explored the interrelationship between ELF lexicogrammar and pragmatics. It was found pragmatic motives often result in lexiogrammatical changes, and vice versa lexicogrammatical innovations have influence on pragmatic norms and strategies. This is evident in the case of third person singular zero by ELF speakers for the sake of efficiency of communication.

Although these studies have unveiled much about systematic linguistic practices involved in lingua franca communication, and most have concentrated on spoken data from established ELF corpora, only three (e.g., Björkman, 2008a, 2010, 2011), to my knowledge, dealt solely with academic lectures, the description of which is expected to provide crucial information about English-medium education.

Björkman (2008a) investigated the morphosyntactic features of spoken ELF

by engineering students and lecturers in a bilingual university in Sweden. The data of the study consists of 42 hours of 'monologic' speech, i.e., lectures, and 28 hours of 'dialogic' speech, i.e., student group-work. The study reports relatively few cases of non-standard language use at the morphological level in both speech events. However, the diverged language use in terms of syntax is much more frequent and varies between the two speech events, especially at the clause level. While there are many incomplete sentences in dialogic speech, monologic speech mostly consists of much longer and complete sentences. In addition, 'non-standard question formulation' and 'negation' occur more often in student group-work whereas pre- and post- dislocations are much more frequent in lectures. It was found that few disturbances in communication occur at the morphosyntactic level since engineering personnel seem to favor function and intelligibility over redundant language features and accuracy.

Björkman (2010) analyzed 21 ELF as the Medium of Instruction lectures (42 hours and 44 minutes) from 13 different lecturers at a large technical university in Sweden. After extensive analyses of non-native like morphosyntactic uses of English, four lectures were fully transcribed for the investigation of pragmatic strategies used by the lecturers. The study reported seven pragmatic strategies adopted in the lectures, including *commenting on terms and concepts*, *commenting on the content of task*, *commenting on discourse structure*, *commenting on intent*, *signaling importance*, *repetition*, and *questions*. However, it was found lecturers in the study generally make little use of pragmatic strategies in comparison with students in the ELF settings.

Therefore, the author suggested that lecturers increase lecture interactivity and create more opportunities for the use of pragmatic strategies so as to prevent misunderstanding and boost comprehension, which has been proven effective in dialogic speech events of students.

Originating from author's doctoral project as well as Björkman (2008a, 2010), Björkman (2011) investigated pragmatic strategies in lectures and student group-work sessions in the ELF as the medium of instruction context at a technical university in Sweden. The study transcribed the dataset selectively for extensive identification of non-standard usage, and fully transcribed four lectures and four student group-work sessions for intensive analysis. The results show that while speakers in both speech events employ a variety of pragmatic strategies, lecturers used fewer pragmatic strategies than students did in group-work sessions. Despite varying levels of frequency among the students, their richer use of pragmatic strategies could be related to the nature of group-work speech event they are involved in. Considering the critical role of pragmatic strategies, the study recommends lecturers in ELF settings create more opportunities to use pragmatic strategies for the purpose of communicative effectiveness.

2.5.3 Critique of previous studies on academic lectures in English as a lingua franca settings

These studies made rich description of practices involved in ELF communication, which is illuminating for educational practitioners, students and

anyone who might be communicating in ELF settings. It was found non-native-like morphosyntactic language use rarely causes communication breakdowns and most deviations from native pragmatic norms, except for non-standard question formulation (Björkman, 2008a), do not seem to lead to disturbance or misunderstanding. These studies (e.g., Björkman, 2010, 2011) strongly recommend lecturers employ more pragmatic strategies for the purpose of communicative effectiveness. The problem seems to lie in how to linguistically implement these pragmatic strategies instead of whether to use them. Since non-native-like morphological use usually do not cause miscommunication, it is advisable to investigate ELF language use beyond the word level. Therefore, formulaic language of word sequences, which allows both linguistic convention and creativity (Mauranen, 2009), makes practical and reasonable research object in ELF studies. Meanwhile, rhetorical structures at the discoursal level across cultures also merit exploration.

2.6 The proposed study

On the basis of the literature reviewed, a few gaps could be discerned.

Firstly, genre analysis of university lectures and the study of MWEs in lectures have been generally lacking. Methodologically, mixed methods that combine the advantage of genre studies and the power of corpus linguistics are relatively few.

Secondly, the majority of previous studies on university lectures has been confined to native English countries, with those in non-native English countries overlooked.

Relatively few studies have been conducted across cultures or in ELF settings. Given the unprecedented upsurge of globalization in education, there is urgent need to compare the genre conceptualization and formulaic language use between native and non-native English lecturers.

Thirdly, the frequency-based criteria for the identification of lexical bundles are problematic in terms of methodology and pedagogy. Other dimensions such as structural-semantic unity, multi-functionality, and formal continuity/discontinuity also deserve attention.

Fourthly, most research on bundle-move connections were in misalignment of the defining features of the ESP approach of genre analysis in that they adopted the bottom-up approach to explore such connections. They inappropriately took textual proximity for bundle-move connections. Other studies (e.g., Mizumoto, Hamatani & Imao, 2017) on bundle-move connections are statistical in nature, and need to be triangulated and corroborated by studies that address the bundle-move connection problem directly from the perspective of communicative functions.

In view of the gaps identified above, it would be of theoretical and pedagogical value to conduct a contrastive study of university EMI lectures given by native and non-native English lecturers. Specifically, the proposed study will: 1) identify and compare the rhetorical move structure of university EMI lectures given by native and non-native English lecturers; 2) specify and compare linguistic features related to classroom discourse of university EMI lectures given by native and non-native English lecturers,

with particular attention being given to formulaic language; 3) situate formulaic sequences in their rhetorical move structure and expound how formulaic language may help achieve the communicative purposes of the corresponding moves/steps.

2.7 Summary

This chapter reviews literature related to the proposed study. Specifically, it sets the ESP school of genre analysis as the theoretical framework for the investigation of university EMI lectures. Then, it presents relevant studies on linguistic features of academic lectures. It also introduces the concept of English as a lingua franca to discuss language practices in intercultural communication, academic lectures in particular. The details of research methodology will be discussed in Chapter 3.
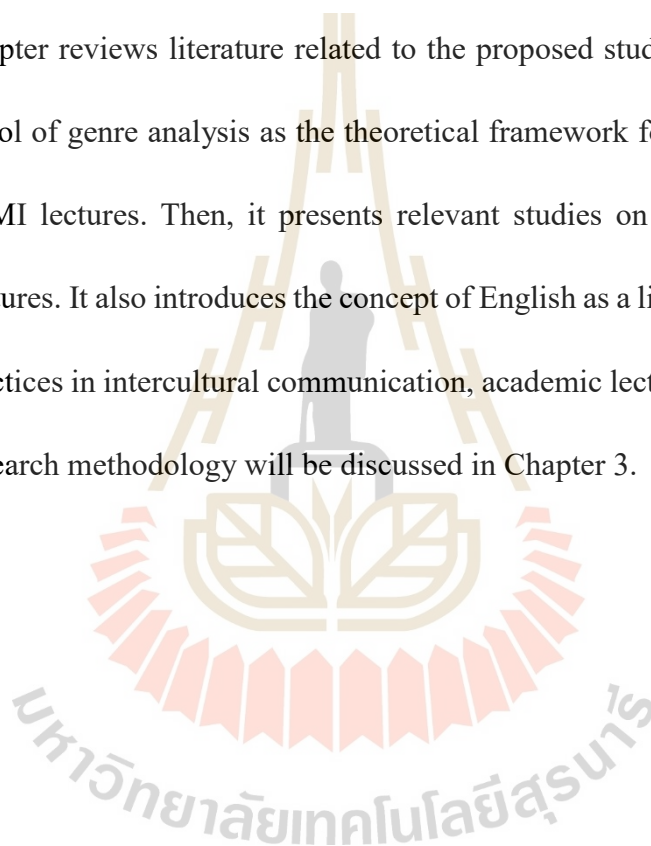
CHAPTER 3

RESEARCH METHODOLOGY

This chapter describes methodology employed in the present study. First, it presents a brief account of the research design. Then, it specifies the principles and procedures of the construction of corpora to be used in the study. Following that, the identification and the comparison of rhetorical move structure and formulaic language in CCL and CEL lectures, as well as the move-formulaicity connections, are discussed. The next part reports the pilot study, and its initial findings and major revisions. The last part ends the chapter with a summary.

3.1 Research design

To address the problems concerning the EMI practices in China and answer the corresponding research questions, a three-phase mixed method research was designed. In the first phase, a contrastive genre analysis was conducted on move structure between university EMI lectures in CCL and CEL. The research results in this phase are expected to answer research question 1). The second phase compared the use of formulaic language between the two corpora, the results of which will provide answers to research question 2). The third phase examined formulaic sequences in relation to the corresponding moves/steps. The results in this phase answer research question 3).

Fig 3.1 illustrates the major elements of the research design and its essential procedures.

Research questions
1. Rhetorical move structure
2. Formulaic language
3. Move-Formulaicity connections

Research design
1. Genre analysis and comparisons
2. Formulaic language comparisons
3. Move-formulaicity connections
    and comparisons

Corpora construction
1. CEL
2. CCL

Phase I: Genre analysis
1. Move identification and coding
2. Rhetorical structure for EMI classes
3. Move Sub-corpora
4. Comparing moves/steps between CCL and CEL

Phase II: Formulaicity
1. Formulaic sequences identification
2. Comparing formulaic sequences in CCL and CEL in
    terms of frequency and structure

Phase III: Move-formulaicity connections
1. Connecting formulaicity with moves/steps
2. Comparing move-formulaicity connections between
    CCL and CEL

Figure 3.1 Research Design Flow Chart

(CEL= Corpus of English Lecturers; CCL= Corpus of Chinese Lecturers)

3.2 The corpora

The proposed study seeks to investigate the rhetorical move structure as well as formulaic language in EMI lectures given by native and non-native English lecturers, and examine formulaic sequences in relation to the rhetorical moves/steps. In this regard, the extant general corpora may not be suitable for the research objectives, since general corpora are typically designed for the description and/or generalization of overall features of a language or language variety on the basis of frequencies and co-occurrences of certain linguistic items or patterns (e.g., collocation, colligation) (McEnery, Xiao & Yono, 2006; Aston, 2001). The defining features and the large size of general corpora render it inapplicable to refer to any general corpus for the present study. It is therefore essential to construct specialized corpora for the present study, since their manageable size and homogeneous composition make them more appropriate for such qualitative studies as genre analysis.

A specialized corpus is a corpus of texts representative of a specific register, discourse domain, or subject matter (Hunston, 2002; De Beaugrande, 2001). The present study selects university lectures from MICASE and BASE to compile the small lecture corpus given by native English lecturers. Since there is no ready-made corpus for university EMI lectures given by Chinese lecturers, the present study self-built a specialized corpus for the research.

3.2.1 Corpus size

A methodological concern for corpus-based studies is related to the external validity of the corpus in question, whereby it is essential to ensure the representativeness of corpus involved so that it may truly represent the discourse domain under investigation (Biber, Connor & Upton, 2007, p.17). This gives rise to issues concerning the size and representativeness of the corpus.

The size and representativeness of the corpora are generally considered thorny issues in corpus studies with no easy answer. Some insist that the representativeness of a specialized corpus be measured by the degree of 'closure' (McEnery & Wilson, 2001) or 'saturation' (Belica, 1996) of the corpus. The concept of 'closure/saturation' refers to the fact that certain linguistic features of a discourse domain/genre appear to be stable or show very limited variation beyond a certain point (McEnery, Xiao & Tono, 2006). However, the saturation of a corpus is only concerned with saturation at the lexical feature level while few attempts have been made to represent saturation of other linguistic features such as part of speech, and sentence types, let alone rhetorical move structure of texts.

For many other corpus researchers, there might be no ideal size for a corpus since it generally depends on the research needs and purposes (Flowerdew, L., 2004). To a large extent, the optimum size of a corpus is related to the linguistic features under discussion. If the target feature to be examined occurs quite frequently (e.g., the use of nouns, verbs, personal pronouns, etc.), the corpus can be smaller; however, if the target

feature is less common, it is advisable to investigate it in a larger corpus.

Despite what has been discussed above, a specialized corpus should be of adequate size for particular features to recur frequently enough to reveal the patterns of use. In this case, previous studies may be good references for the design of corpus size.

Another caveat to bear in mind is that a specialized corpus is not to be confused with a small corpus, although a specialized corpus tends to be small in size. The size of a corpus is a relative and evolving concept. A large-scale corpus of the 1960's, e.g., the one-million-word Brown corpus of written American English and its British counterpart the one-million-word Lancaster-Oslo-Bergen (LOB) corpus, is extremely small today given the emergence of the new giga-level online corpora, e.g., the 14-billion-word Intelligent Web-based Corpus and the 6.04-billion-word News on the Web (NOW) Corpus, etc.

3.2.2 Corpus of university EMI lectures given by native English lecturers

The Michigan Corpus of Academic Spoken English (MICASE) is an academic spoken language corpus of contemporary university speech collected within the microcosm of the University of Michigan in Ann Arbor between 1997 and 2002. It contains data from a wide range of speech events in the university. There are 152 speech events in total in MICASE, 62 of which are lectures. Speech events are also categorized according to other contextual attributes, e.g., discipline, participant level, native speaker status, primary discourse mode (i.e., to what extent a speech event is monologic or interactive), and speech event type (i.e., large lecture vs. small lecture) (Simpson, Lee

& Leicher, 2007).

The BASE corpus is a collection of academic spoken data recorded at the Universities of Warwick and Reading between 1998 and 2005. It consists of 160 lectures and 39 seminars collected in a variety of university departments across four broad disciplines, including Arts and Humanities, Life and Medical Sciences, Physical Sciences, Social Studies and Sciences.

MICASE and BASE have been widely used in many corpus-based studies of academic spoken data, generating many illuminating findings. The comprehensiveness of the corpora makes them good resources to investigate native academic spoken genre. A series of selection criteria have been adopted to select small lectures from MICASE and BASE to build a specialized corpus of university EMI lectures given by native English lecturers, including discipline, primary discourse mode (i.e., the predominant types of discourse characterizing the speech event), speech event type (i.e., large lecture vs. small lecture), interactivity level and native speaker status.

Only science-related lectures are selected to ensure the homogeneity of lectures under discussion. Since the proposed study pays special attention to challenges lecturers face in the delivery of classroom content, monologic lectures, where one speaker (the lecturer) monopolizes the floor of the speech, are selected. The interactivity level is rated on the basis of the average number of words per turn (WPT), whereby 'mostly monologic' is stipulated at 100-450 WPT and 'highly monologic' at 500+ WPT (Simpson-Vlach & Leicher, 2006). It must be noted that students'

contribution in class is essential to effective lecture construction, thus is included in the Phase-I genre analysis of lectures. However, students' language in class, e.g., students' response to lecturers' questions, was not considered when analyzing formulaic language in Phase II and III. In addition, the legitimate class size in Chinese universities is typically 30 students and the real EMI class size in the Chinese universities surveyed is generally less than 30 students. For reference concerns, small lectures with 40 or fewer students from MICASE and BASE are chosen. Lastly, the present study is committed to a contrastive study of EMI lectures given by native and non-native English speakers, therefore, the selection of MICASE and BASE lectures are only confined to those given by native speakers of English.

A retrieval of MICASE on the criterion set of 'mostly monologic', 'small lecture' and 'native speaker' produced a return of a total of 12 hits, six of which are science-oriented, covering biology, engineering, and computer programing. The same criteria were applied to select another six science lectures from the BASE corpus. These twelve small lectures (See APPENDIX A) were renamed as CEL01-12, and put in one folder to form the specialized corpus of university EMI lectures given by native English lecturers (CEL).

3.2.3 Corpus of university EMI lectures given by native Chinese lecturers

As has been mentioned above, there is no ready-made corpus for university EMI lectures given by Chinese lecturers, a specialized corpus (See APPENDIX B) was built for the present study.

The lecture data were collected at a polytechnic university in the northwestern province of Shaanxi, China. The university was chosen because it represents common universities in China in terms of the comprehensive strength. In 2017, the university ranked 329th among the 817 four-year public universities in China (MOE, 2017).

To ensure comparability, the same selection criteria (except for native speaker status) as those for CEL were used for the compilation of the corpus of university EMI lectures given by non-native English lecturers, in this case, the Chinese lecturers. Therefore, only those monologic lectures (*post hoc* analysis of lectures collected all reach the level of 100+ WPT, thus all considered monologic) with no more than 40 students given by Chinese lecturers are candidates for the target corpus.

The lectures used for the construction of CCL come from the classroom teaching videos accessible from the university website. The university website offers links to five different EMI courses for international students, three of which include the real teaching videos shot between 2016 and 2018. These three courses are *An Introduction to Earth Sciences*, *Recent Advances in Oil and Gas Production Engineering*, and *Advanced Mathematics I*. There are altogether 14 EMI lectures downloadable from the university website, among which ten were chosen according to the general level of interactivity and lecturer. Although the website of a fourth course *Oil Drilling Engineering* did not provide much content of the course itself, including the classroom teaching video, the researcher managed to get the video clips from the

Information Center of the university with the permission of the lecturers involved. Two out of the three lectures were selected for the study. It turned out all these lectures are mostly monologic, with teachers taking the major role of instructor in the classroom, and all are in small size, with 15 to 20 students.

All these teaching videos were copied and stored in a computer for transcription. For the concern of easy readability and comparability, the present study followed the MICASE orthographic transcription conventions and mark-up system. The paralinguistic information such as the tones and gestures, though important for the description and comprehension of the speech event, is not the concern of the present study, therefore was not transcribed. Each lecture sound file was transcribed in the text format and renamed as CCL01-12.

With regard to the external validity and representativeness of the corpora, previous studies might be used for reference. When examining the rhetorical move structure of second language classroom discourse (L2CD), Lee (2011) and Lee (2016) used a self-created corpus with 24 lessons from four highly experienced EAP teachers. The corpus has a total number of 179,738 running words. In another comparative study of EAP lessons and university lectures, Lee and Subtirelu (2015) selected 18 EAP lessons from corpus of Lee (2011) and built a L2CD corpus with a total of 105,708 running words. What's more, in summarizing defining parameters for specialized corpora, Flowerdew, L. (2004) proposed that the size for a specialized corpus be in a range of 20,000–250,000 words.

Given the data available and the previous studies, the corpus size should serve the purpose of the research. Although the data are not large enough to represent all universities and colleges in China, they may provide authentic scenarios where some typical features as to the organization of lectures and the common formulaic language Chinese lecturers often use in EMI classes could be described.

3.3 Rhetorical move structure

According to Biber, Connor and Upton (2007), fundamental to corpus-based discourse studies is to "segment a text into higher-level structural units" (p. 1), well-defined discourse units beyond the sentence level. After the compilation of the CEL and CCL, their rhetorical move structure was analyzed for comparison.

3.3.1 Analytic approach

Generally, two approaches have been employed in previous research in the segmentation of the discourse units: top-down and bottom-up approaches.

The major difference between the two approaches lies in the analyses conducted to the discourse, viz., functional versus linguistic analyses. The top-down approach starts off with determining a set of discourse units (i.e., the 'moves') according to the major communicative functions they serve in the discourse. Then linguistic analyses are conducted afterwards to explicate how these functionally-defined discourse units are linguistically realized. Functional investigation is primary while linguistic description is secondary.

In contrast, the bottom-up approach begins with the linguistic description of the discourse (e.g., the 'Vocabulary-Based Discourse Unit' (VBDU) analysis (Csomay, Jones & Keck, 2007)), automatic segmentation of texts into VBDUs on the basis of vocabulary patterns as well as linguistic grouping of VBDUs in accordance with their lexico-grammatical features. Then functional analyses are conducted to regroup the VBDUs. To a great extent, functional analyses mainly serve an interpretative role to investigate the systematic functional characteristics of the linguistically-defined discourse units. Contrary to top-down approach, linguistic description is primary while functional analysis is secondary in the bottom-up approach.

The use of linguistic criteria in the bottom-up approaches contradicts the genre conceptualization of Swales (1990) that sees communicative purpose/function as a privileged and practical criterion for genre identification. In fact, move identification (textual division/staging) involves cognitive inquiry of functions the text segments serve with regard to convention, appropriacy, and content rather than searching for linguistically defined boundaries (Paltridge, 1994; Bhatia, 1993). Therefore, the present research adopts a top-down approach in move/step identification.

3.3.2 Coding scheme

Among the few genre studies of classroom discourse, Young (1994) adopted the SFL perspective in the study of macro-structure and prominent micro-features of university lectures. This is in misalignment with the theoretical framework of the ESP approach adopted in the present study. Other research adopted the Swalesian framework

in their genre analysis of classroom discourse, but they are confined to individual lecture sections/phases instead of the whole lecture discourse (e.g., Thompson (1994) and Lee (2009) on lecture introductions, Cheng (2012) on lecture closings).

Nevertheless, Lee (2011) followed Swales' (1990) move analysis framework, and analyzed the rhetorical move structure in the L2CD corpus. Extending the work of Lee (2011), Lee (2016) adopted a combined methodology of Swalesian move analysis and corpus-based techniques to explore the rhetorical move structure and linguistic features of classroom discourse of an EAP program.

Theoretically, Lee (2016) adopted the Swalesian genre analysis framework, a robust framework for researching specialized discourses, which is also the theoretical orientation of the present study. In addition, the rhetorical move structure and linguistic features in Lee's (2016) study are also the content of the present research. Besides, similar to Lee's (2016) research, the present study also intends to investigate the whole classroom sessions. Taking all these commonalities into account, the present study attempts to refer to Lee's (2016) move structure as its framework.

It must be noted that Lee's (2016) research is an examination of language classes, whose purposes are to facilitate language learning through classroom activities such as repetition and substitution drills, pair work, role playing and games. University lectures, however, are more concerned with knowledge transmission and clarification of research procedures and events. Van Lier (1988) identified four types of L2 classroom interaction on the bases of both activity-orientation and topic-orientation.

Though the boundaries between different classroom interaction types are often not clear-cut as a result of the flexible use of activity talk and topic talk by lecturers, different courses do favor different classroom interaction types. Language classes are often less topic-oriented and more activity-oriented, whereas university EMI lectures tend to be more topic-oriented and less activity-oriented. Therefore, the framework of Lee (2016) needs to be fine-tuned to serve the research object of university lectures.

An initial rhetorical move structure scheme was thus proposed on the basis of Lee (2016) and initial reading of three lectures from both CEL and CCL. The scheme was then revised reiteratively by the researcher and negotiated with an inter-coder. The final coding scheme is illustrated in Figure 3.2.

Opening Phase

M1: Getting Started
M2: Warming Up
M3: Setting Up Lecture Agenda

Theme Network Building Phase

M4: Introducing The Topic
M5: Elaborating The topic
M6: Building Theme Network
M7: Making Aside
M8: Housekeeping
M9: Checking Comprehension & Consolidating Learning

Closing Phase

M10: Wrapping Up Lecture
M11: Cooling Down
M12: Ending Lecture
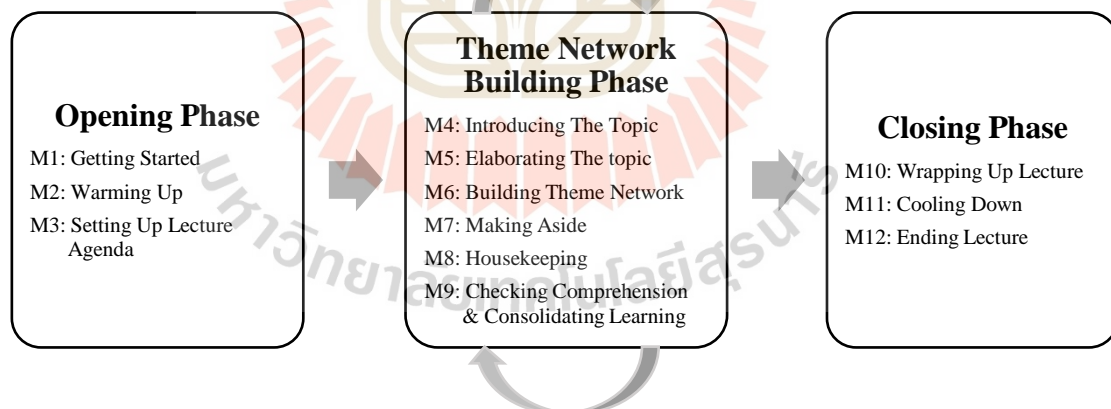
Figure 3.2 Move coding scheme of university EMI lectures

The major difference between the move structure of language classes and university EMI lectures lies in the second phase. The *activity cycle phase* in language classes (Lee, 2016) typically consists of three moves, viz., *M4: Setting Up Activity*

Framework, *M5: Putting Activity In Context*, and *M6: Reviewing Activity*. However, university EMI lectures are usually organized with reference to knowledge topics closely-related to the lecture theme. Lecturers' priority is often given to building theme networks to impart knowledge to students, with the *Theme Network Building Phase* as the main part of university EMI lectures. Initial reading of three EMI lectures from CCL and CEL identified six moves in the second phase of lectures, including *M4 Introducing The Topic*, *M5 Elaborating The Topic*, *M6 Building Theme Network*, *M7 Making Aside*, *M8 Housekeeping*, and *M9 Checking Comprehension & Consolidating Learning*.

3.3.3 Coding process

The present study adopts a top-down approach to move analysis, which is, to a large extent, determined by the communicative purposes the writer/speaker intends to accomplish. That involves much of the human cognition process. In addition, depending on the communicative functions the speaker intends to achieve, the move lengths may vary from several sentences to one sentence or even a few words, resulting in multi-functional sentences, whereby more than one communicative purposes appear to be realized within that sentence. In case of such move embedding in one sentence, the sentence was identified and coded as one single move/step with reference to most dominant communicative purpose (Holmes, 1997; Ozturk, 2007; Hirano, 2009).

More topic-oriented, lecturers are more concerned with knowledge dissemination, which could be centered upon the linguistic expression of

communicative functions. These concerns may be articulated through various linguistic patterns, including the use of formulaic language, at the step level to express specific communicative functions. In alignment of studies in RA rhetoric structure (e.g., Yang & Allison, 2003; Amnuai & Wannaruk, 2013; Cotos et al., 2017), the researcher considers the step as "more appropriate level for investigating the function-form gap than is the move" (Moreno & Swales, 2018) in academic lectures. Therefore, the present study codes university EMI lectures and calculates inter-coder reliability at the step level.

The present study uses BFSU Qualitative Coder 1.1 (Xu & Jia, 2011), a powerful software program for qualitative research, for the coding of moves and steps. This tool allows users to customize codes to be assigned to specified texts (see APPENDIX C for a sample of move-coded text and the operating window of BFSU Qualitative Coder 1.1), and presents immediate statistical data of the qualitatively coded texts (see APPENDIX D for a sample of move statistical data), thus suits the present research.

3.3.4 Inter-coder reliability

Though a move/step could be identified according to the salient functions, it unavoidably may involve certain degrees of subjectivity (Holmes, 1997, p. 325). In order to tackle the subjectivity problem, the present study adopts inter-coder reliability checks. The coding of moves/steps involves two independent coders, including the researcher and a doctoral student in the field of applied linguistics. The researcher

randomly selected three lecture transcripts (25%) from both CCL and CEL, and tried to

analyze them with reference to the coding scheme adapted on the basis of the move

types established by Lee (2016). Move/step types were adapted and new move/step

types were added to create a draft of coding scheme. Then the researcher explained in

detail the coding scheme to the second coder. After that, the second coder used the

coding scheme to analyze the very same six lectures the author had analyzed. Any

disagreement of the move/step assignment as well as inclusion and exclusion of

move/step types were discussed with a third coder, an expert in the area of genre

analysis, until a consensus was reached. In addition, definitions/descriptions of each

move and step were fine-tuned, and examples of move/step types were given. Finally,

a coding protocol was created to present an inventory of move/step types, as shown in

table 3.1.

Table 3.1 Move structure coding protocol for university EMI lectures

Phases	Moves/step	Descriptions
Opening Phase	*M1 Getting Started*	signals the official start of a lecture.
	M2 Warming Up	leads students to the present lecture through presenting lead-in information, recalling previous lecture, looking ahead to future lecture, and highlighting course-related issues before engaging in more substantive parts of the lecture.
	M2S1 Leading in	prepares students for the lecture with issues directly or indirectly related to the lecture.
	M2S2 Recalling previous lecture(s)	presents contents of previous lectures to help refresh students' memory so as to prepare them for the present lecture.
	M2S3 Looking ahead	informs students of the upcoming lecture.

Table 3.1 Move structure coding protocol for university EMI lectures (Cont.)

Phases	Moves/step	Descriptions
Opening Phase	*M2S4 Housekeeping*	is used to make class management announcements, and offer reminders.
	M3 Setting Up Lecture Agenda	briefs contents to be covered in the present lecture.
	M3S1 Announcing/ Clarifying lecture theme	explicitly introduces the lecture theme.
	M3S2 Providing lecture scope	enumerates contents to be covered in the lecture.
Theme Network Building Phase	*M4 Introducing The Topic*	introduces topics germane to the lecture theme.
	M5 Elaborating The Topic	concerns with elaborating topics involved.
	M5S1 Explaining terms	gives detailed explanation of technical terms involved in the topic.
	M5S2 Highlighting importance	highlights importance of the topic.
	M5S3a Presenting background knowledge	directly presents background knowledge of the topic.
	M5S3b Contextualizing real-word experience	involves referring to the real-word experience of students as the background to help understand the topic.
	M5S3c Setting up premise	presents known facts or conditions as the foundation for understanding a certain topic.
	M5S4 Expounding rationale	elaborates on or rephrases the theories, mechanisms, motivations, as well as common practices/knowledge related to the topic.
	M5S5 Demonstrating the topic	involves explaining in detail the factual information, methods, processes, procedures, and results related to the topic under discussion, often with examples and data.
	M5S6 Providing caveats	explicitly articulates caveats related to the topic.
	M5S7 Making comments	allows lecturers to share with students their understanding of and comments on the topic.
	M5S8 Summing up the topic	summarizes the topic to consolidate learning.
	M5S9 Initiating co-building lecture	engages students in co-building the lecture, thus facilitating learning
	M5S10 Pinning down and/or clarifying the topic	confirms and clarifies the topic under discussion.

Table 3.1 Move structure coding protocol for university EMI lectures (Cont.)

Phases	Moves/step	Descriptions
Theme Network Building Phase	*M6 Building Theme Network*	helps build a whole knowledge network that may serve the lecture theme.
	M6S1 Referring to previous lecture/other source	involves creating knowledge network by connecting the present lecture with previous lecture (parts) or knowledge from other sources.
	M6S2 Specifying subsequent/future content	reserves a node in the knowledge network by specifying subsequent/future content that may support the theme in the present lecture.
	M6S3 Connecting/Comparing topics	connects or compares topics.
	M7 Making Aside	refers to classroom talk less germane to lecture content, but may be conducive to understanding the topic, as well as building and maintaining good teacher-student rapport.
	M8 Housekeeping	is used to give recommendations, and offer reminders.
	M9 Checking Comprehension & Consolidating Learning	serves to check on students' comprehension as well as to consolidate learning.
Closing Phase	*M10 Wrapping Up Lecture*	makes summaries of the lecture content.
	M11 Cooling Down	serves to attend to course-related matters and/or to discuss future lectures.
	M11S1 Housekeeping	is used to assign homework, and offer homework-related reminders.
	M11S2 Looking ahead	presents students a preview of upcoming lectures.
	M12 Ending Lecture	declares the ending of the lecture.
	M12S1 Dismissing class	signals the ending of the lecture.
	M12S2 Farewell	ends the lecture with farewell and/or gratitude.

Utilizing the coding protocol, the two coders independently coded another three lecture transcripts (25%) from CCL and CEL to check inter-rater reliability in move identification, the results of which turned out to be high, 95.3% agreement between the coders. Then the researcher independently coded the rest of the lecture transcripts.

3.3.5 Move sub-corpora compilation

All move-coded text segments were copied in separate text files to form move sub-corpora. It must be noted that the present study complied sub-corpora at step level instead of move level because steps provide more detailed linguistic realizations of moves, which are of more practical value for the communicative purposes lecturers endeavor to convey. Therefore, individual steps were put together to form individual sub-corpora, though they are still referred to as move sub-corpora for convenience.

Specifically, all the text segments with the same move-step label from the same lecture were copied in one individual text file, and these files were renamed in the format of move-step type label + Lecture number. Texts with the same move-step type label from the 12 lectures were then put in the same folder. Eventually, the folders were renamed with the move-step type label as individual move sub-corpora. Identical procedures were adopted to compile move sub-corpora for both CCL and CEL.

3.3.6 Comparing rhetorical moves in CCL and CEL

When all the moves and steps are coded, their frequency statistics is automatically calculated through BFSU Qualitative Coder 1.1. A customary follow-up step is to determine the conventionality of individual moves/steps. Although different scales of move stability categorization were adopted in previous studies, most favored the conventional vs. optional dichotomy. Nevertheless, the cutoff frequency of move conventionality is always arbitrary. While Nwogu (1997) and Loi & Evans (2010) set it at 50% occurrences for research articles, Kanoksilapatham (2005, 2011) established

60% frequency as the cutoff in her study of research articles and abstracts. Lee (2016) raised the cutoff to 80% due to the small corpus size and the speaker number in the genre analysis of second language classes. However, considering the real-time nature of lectures and relatively larger lecture number, the present study set the cutoff at 50%. A move/step must occur in 50% of all lectures to be categorized as conventional, whereas any move/step with a frequency below 50% is rendered optional. Unlike move analysis for written texts, no attempt was made to determine a sequential order of moves and steps due to the frequent online processing involved in lectures. Then, the rhetorical move structure of university EMI lectures given by native and non-native English lecturers was proposed respectively and compared with each other.

Lorés-Sanz (2016) adopted an ELF perspective and examined rhetorical move structure simplification and hybridization with reference to the variation of move numbers and move structure hybridization such as focalizing move structure, move chain structure and cascading move structure in abstract writing of non-native academics. Focalizing move structure involves detailed description of one move in abstract writing, move chain structure refers to the alternate occurrence of certain moves, whereas in cascading move structure some traditional moves may be embedded in others in the abstracts of non-native academics. The move chain structure and cascading move structure are akin to move cyclicity where certain moves may have closer affinity to each other, and have the tendency to recur. This has been researched in the study of research articles by many researchers (e.g., Amnuai & Wannaruk, 2013;

Peacock, 2002; Posteguillo, 1999; Stoller & Robinson, 2013; Yang & Allison, 2003).

Essentially, statistical and textual analyses were conducted to see if rhetorical move structure simplification and/or hybridization occur in the CCL lectures. It might be of practical value to compare the move cycles of EMI lectures by native and non-native English lecturers since they promise to reveal more detailed and clearer instructional schemata of lecturers across cultures. In the present study, if a cluster of moves recurs six times in a cyclical pattern in a range of six lectures, it is considered a move cycle.

On the basis of text analysis results, semi-structured interviews (See APPENDIX E guideline interview questions 1-14) were designed and conducted with four Chinese EMI lecturers (25%) concerning their conceptualization on rhetorical move structure. All the interviews were fully transcribed, but only selectively translated. The general information of the interviews is given in table 3.2.

Table 3.2 General information of semi-structured interviews

Interviewee	Professional title	English speaking country experience	Education background	EMI teaching experience	Interview duration
Interviewee A	Professor	18 months	PhD	5 years	23 minutes
Interviewee B	Lecturer	8 months	Master	12 years	24 minutes
Interviewee C	Associate Professor	12 months	PhD	10 years	38 minutes
Interviewee D	Associate Professor	12 months	PhD	5 years	22 minutes

3.4 Formulaic language

As is the common practice of top-down move analysis, when all the moves are identified, it is essential to explore how the rhetorical/communicative functions are linguistically realized. According to Swales (1990), the communicative functions of a genre exhibit typical verbalization patterns recognizable by members of the discourse community. In the present study, linguistic features as instantiated by formulaic sequences in the move sub-corpora are identified, analyzed, and compared between native and non-native English EMI lecturers.

3.4.1 Identification of formulaic sequences

It must be noted that formulaic sequences can be continuous and discontinuous in accordance with their fixedness, ranging from fixed idioms (e.g., *kick the bucket*, *carry coals to Newcastle*) and invariable expressions as a result of grammticalization or lexicalization processes (e.g., *above and beyond*, *of course*), to expressions that allow a certain degree of compositional variation and semantic transparency (e.g., *play a vital/important/major role*), and even frame patterns with open slots to be filled (e.g., *as X as possible*, *too X to*).

The identification of formulaic language has never been easy due to its multi-faceted quality and inherent complexity. Nonetheless, the study intends to be as inclusive as possible for the sake of the maximum pedagogical value of formulaic language. Therefore, the present study adopted mixed criteria that involve a series of diagnostics adapted from Wray (2008) and Namba (2008) to identify as many FSs as

possible since "most examples will be captured one way or another" (Wray, 2008, p. 110).

Eleven diagnostic criteria for FSs were proposed, tested, and fine-tuned by Wray (2008) on the basis of feedbacks from a group of eight native English-speaking PhD students and post-doctoral researchers, and nine high proficiency non-native English speaker studying for doctoral or post-doctoral programs. They include: A) Grammatical irregularity; B) Semantic opacity; C) Situation/register/genre specificity; D) Pragmatic function; E) Idiolect; F) Performance indication; G) Grammatical/lexical indication; H) Previous encounter; I) Derivation; J) Inappropriate application; and K) Mismatch with maturation.

However, the present study focuses on EMI classroom language by native and non-native English lecturers, which is different from Wray (2008), who examined the language of English L1 children, L2 children, L2 teenagers, L2 adults, as well as L1 aphasic patients. Therefore, the diagnostic criteria of Wray (2008) were reshuffled and adapted to cater for the research purpose of the present study.

Firstly, the present study deals with the mundane and common EMI lectures, where situation specific FSs were rarely used except for the classroom citation by lecturers to explain religious rituals, so criterion C) Situation/register/genre specificity was not included as a criterion. Secondly, criteria D) Pragmatic function was also excluded since it is mainly concerned with the social consequences or pragmatic functions of FS language, which deviated considerably from the knowledge

dissemination purposes of EMI lectures. The rest nine criteria, together with criterion diagnostics L) underlying frame by Namba (2008) were regrouped into the following operational criteria for purpose of the study.

1) Grammatical irregularity and/or semantic opacity

It was suggested that ungrammaticality found in some FSs might be attributed to the fossilization of certain MWE forms, which were initially processed holistically. Meanwhile, grammatical irregularity may, to a great extent, presuppose semantic opacity. When a MWE is grammatically irregular or ungrammatical in the traditional sense, it tends to be semantically opaque. Therefore, these two criteria were combined as one. This criterion stipulates that a MWE is considered to be a formulaic sequence as long as it is not strictly predictable from its component parts or grammar whether in form or meaning (e.g., *suffice it to say*, *play a vital role in*, *shed light on*, *above and beyond*).

2) Morpheme Equivalent Unit (MEU) rule

It is worth noting that the frequent ambiguity in terms of grammatical irregularity and/or semantic opacity often jeopardize the effectiveness of the first rule, leading to subjective evaluation and judgment of expressions such as *at this time*. Superficially, even a learner who encounters *at this time* for the first time stands a good chance of understanding the expression simply by adding the individual components *at + this + time*, thereby it may be treated as compositional and non-idiomatic, non-FS. Nevertheless, "[f]ull compositionality is rarely the case" (Taylor, 2006, p. 61), and even

the simplest collocations may pose comprehension difficulty for learners, which may create a chance for learners to process them holistically, i.e., treating such word strings as whole units. This is where the MEU came into play, whereby '*at this time*' is considered as a formulaic sequence though it does not strictly meet the semantic opacity criterion. MEU, heteromorphic in itself, may refer to a word or word string processed like a morpheme without recourse to any form-meaning matching of its sub-parts (Wray, 2008). A MEU-like word string is considered a formulaic sequence.

Though native English dictionaries such as *Oxford Advanced Learner's Dictionary* and *Collins COBUILD Advanced Dictionary*, and phrasal expression list of Martinez and Schmitt (2012) were constantly consulted to decide the FS status of MWEs in the native English lecturers' corpus (CEL), the application of the 'MEU rule' is of essential importance in identifying FSs in its non-native counterpart of CCL since the deviated or approximated forms of FSs have never been listed in any native-English dictionary entries or native-English-oriented FS studies.

3) Formal approximation and functional equivalence

Criteria E)-K) are, more or less, related to the imperfect FS use by the language users due to various reasons such as idiosyncrasy, immaturity, and cerebral damage. These criteria are particularly useful in examining the FS use in ELF contexts. Deviations from standard FS usage (e.g., *in other word* instead of *in other words*, *on conclusion* instead of *in conclusion*) were also considered formulaic when they have corresponding standard formed structures and serve similar functions in the context.

This criterion might be, to certain extent, dependent upon the first criterion, but they are presented as a separate diagnostic in light of their immense role in the identifying the approximated FSs in ELF settings.

4) Underlying frame

This criterion mainly involves formulaic frames that link parallel structures with items of similar characteristics, e.g., not only…but also, as…as, or those with open slots to be filled, e.g., *in Figure N* (N stands for a number).

In short, a formulaic sequence is a sequence of continuous or discontinuous words that expresses a cohesive meaning or performs a certain holistic function, which is not easily discernible by decoding the individual words alone, and it does not have to strictly stick to the native speaker norm morpho-syntactically.

The identification criteria may overlap or even in conflict with each other, a clear indication of the complex nature of FSs. For example, the first two criteria may seem contradictory on the surface. However, they are formulated from different language learning perspectives. The rule of grammatical irregularity and/or semantic opacity originates from phraseology whereas the MEU criterion is concerned with the psychological processing and storage of lexicons. They happen to complement each other in identifying formulaic language. The guiding principle for all these criteria is to identify as many FSs as possible. A MWE has to meet at least one of the criteria above in order to be identified as a FS, though in the actual sense many may satisfy more than one criterion.

3.4.2 Structural characteristics of formulaic sequences

The identified formulaic sequences were then annotated with the structural patterns based on the categories differentiated in Biber et al. (2004) and Wang (2017). Their frameworks were used for reference because they both examined academic spoken data in university teaching registers, similar in nature to the data in the present study. Biber et al. (2004) distinguished three main types of lexical bundles, including type 1) verb phrase (VP) fragments, type 2) dependent clause fragments, and type 3) noun phrase (NP) and prepositional phrase (PP) fragments and comparative expressions. On the basis of Biber et al (1999), Wang (2017) identified five main categories of lexical bundles, i.e., NP fragments, PP fragments, VP fragments, clausal fragments, and the Other category.

Some other structural categories, though relatively few in number, were also identified in the present study. Adjective phrase fragments, for example, were added because they are frequently used as complement to denote the character or status of the subject, e.g., *is equal to, are familiar with*, *exactly the same*, though a very limited part of them were also used as modifiers to describe the amount of the head word, e.g., *quite a few*, *only a few*, and *more and more*. Therefore, these two types of AP fragments use were listed separately as AP(C) and AP(M) since they serve different syntactic functions. Adverbial phrase (AdP) fragments such as *first of all*, *as well*, *as closely as possible*, *all of a sudden*, *later on*, and *all right* are generally used by lecturers to indicate the sequences, range, degree, time and mood involved in classroom teaching.

Then, there is the conjunction phrase (ConjP) fragment, e.g., *so that*, *if…then*, *not only…but (also)*, and *so long as*. This category was specifically added because, different from Biber et al. (2004) and Wang (2017), the present study also investigates discontinuous multi-word expressions. The structural categories of FSs in the present study are illustrated as in Table 3.3.

Table 3.3 Structural categories of formulaic sequences

Structural categories	Examples
Noun Phrase fragment	*things like that, my point of view, the amount of*
Prepositional phrase fragment	*in the center of, as a result of, instead of*
Adjective phrase fragment as modifier	*quite a few, only a few, more and more*
Adjective phrase fragment as complement	*is equal to, are familiar with, exactly the same*
Clausal (NP/pronoun + verb/adj; *if-*clauses; anticipatory *it* + verb/adj; *there be; independent clause*, etc.) fragment	*I don't know if, I will show you, it was possible to, there is*
Verb phrase fragment	*has to do with, going to talk about, to start with*
Adverbial phrase fragment	*first of all, and so on, as well, later on*
Conjunction phrase fragment	*so that, if…then, not only…but (also), so long as*

3.4.3 Coding of formulaic sequences and inter-coder reliability

The present study used the software program BFSU Qualitative Coder 1.1 (Xu & Jia, 2011) for the coding of formulaic sequences.

In order to minimize the subjectivity in formulaic sequence identification, the present study also involved an independent coder. First of all, the researcher independently coded formulaic sequences in three CCL lectures (25%) and three CEL lectures (25%) according to the criteria in Section 3.4.1. Then the researcher explained

in detail these operational criteria of FSs to the second coder. After that the researcher presented the formulaic sequence candidate lists to the second coder to check for their formulaic status. The two coders reached a high agreement of 94.3% and 97.8% for formulaic language in CCL and CEL. Then the researcher coded all FSs in the remaining lectures in both CCL and CEL. Structural annotation is generally straightforward and unambiguous, and was therefore done by the researcher only.

3.4.4 Comparing formulaic sequences in CCL and CEL

When the FSs were coded, they were compared in terms of frequency and structure between CCL and CEL. Meanwhile, semi-structured interviews (See APPENDIX E guideline interview questions 15-16) were also designed and conducted with four Chinese EMI lecturers as to their views on the different FS use between CCL and CEL lecturers, as well as their attitudes towards non-standard FS use. Thus, formulaic language was also examined from an ELF perspective.

3.5 Connecting formulaic sequences and rhetorical moves

Unlike previous studies where the functions of lexical bundles were analyzed in isolation, the present study examined the functions of FSs in connection with the moves/steps where they are situated.

3.5.1 Dissecting formulacitiy-move connections

This study identifies rhetorical moves/steps and formulaic language separately, and tries to establish tentative connections between rhetorical move

structure and formulaic language, which, to a great extent, increases the validity of the relations between the two constructs.

The basic assumption is that all linguistic and non-linguistic elements in a move/step may contribute to the realization of communicative purposes of the corresponding move/step, though to different levels. The present study focuses on the linguistic elements, formulaic sequences to be exact, in rhetorical moves/steps.

According to the connections with moves, the identified FSs were classified into three broad categories, including move signaling FSs, move building FSs, and move neutral FSs. Move signaling (MS) FSs are the FSs that explicitly signal the communicative purposes of the moves/steps to which they belong. They verbally signal a straightforward connection with their corresponding moves/steps. Move building (MB) FSs refer to the FSs that help build the propositional meanings essential to the realization of the communicative purposes of moves/steps. They indicate a less direct or indirect connection with their corresponding moves/steps. Move neutral (MN) FSs refer to the FSs that do not dramatically affect either the propositional meanings embedded in the moves/steps, or the communicative purposes of the moves/steps. They exhibit loose relation with any specific moves/steps, but occur across different moves/steps, e.g., *you know*, *that means*, *I think*, *I mean*, *for example*.

It must be noted the present study investigates move-formulaicity connections at the step level. Different steps under the same move specify different strategies of realizing the overall communicative purposes of the move, thus

investigating move-formulaicity connections at the step level are practically more valuable. Another caveat is that FSs could be assigned different functional categories, depending on the moves/steps where they are extracted. For example, *at the beginning* in *M3S2* of excerpt 1) is move signaling, but in *M5S5* of excerpt 2) or *M6S1* of excerpt 3) could be move building.

1) *Er **at the beginning** I will show you my favorite picture. In this picture we can discuss the geological time and the evolution in the same picture.* (CCL01: *M3S2*)

2) *Er and then example, Opening of Atlantic, yeah Opening of Atlantic. In one hundred eighty million years it bigger five thousand kilometer. Er yeah, **at the beginning** er it's it was end of Ordovician, er it was end of Ordovician, the Atlan- Atlantic become very much will become much bigger, bigger, than used to be.* (CCL01: *M5S5*)

3) *i should have should have said **at the beginning** [0.3] that some of the material that i'm presenting now obviously is related [0.3] to what [0.4] i presented to you in the second year so some of it will be familiar i hope it'll be familiar [0.4] and you should also look at your second year notes [5.3]* (CEL07: *M6S1*)

As for the criticism of circular reasoning (Pho, 2008) on bundle-move connections, Moreno and Swales (2018) deem it as a key element in hermeneutic methods, which is "a dialectical tacking between parts which comprise the whole and

the whole which motivates the parts, in such a way as to bring parts and the whole simultaneously into view" (Geertz, 1980, p.103). The combined approach also see support from Flowerdew, J. (2002), who admitted adopting simultaneous identification of communicative purpose(s), schematic structure, grammatical features, lexical features, etc.

3.5.2 Coding of formulacitiy-move connections and Inter-coder reliability

All the formulaic sequences were coded with functional categories in connection with moves/steps through software program BFSU Qualitative Coder 1.1 (Xu & Jia, 2011). Then the researcher exported the results to excel files for the second coder to judge the move-formulaicity connections. Their agreement rate reached as high as 92.3%, the results of which were considered reliable.

3.5.3 Comparing move-formulaicity connections in CCL and CEL

When the move-formulaicity connections were coded, the statistical calculation of their functional distribution were conducted and the results were compared between CCL and CEL. In particular, move signaling formulaic sequences were demonstrated in their immediate moves/steps. Each move/step were provided with a list of MS FSs used in CCL and CEL, which might be of great pedagogical value for pre-service, in-service novice teachers, and students.

3.6 Pilot study

In order to help advance the main study, a pilot study was conducted on six randomly selected EMI lectures, with three of each from both CCL and CEL.

3.6.1 Results of the pilot study

The pilot study adopted the Swalesian top-down genre analysis, where moves and steps are generally and ultimately determined by the communicative purposes lecturers intend to convey. After repeated reading of the lecture transcripts, the researcher was able to formulate a tentative scheme of move and step codes with reference to the framework of Lee (2016).

The log-likelihood comparison (log-likelihood=4.32, p=0.038<0.05) (Liang, Li & Xu, 2010) of rhetorical move structure indicates significantly more moves/steps used by EMI lecturers in CCL than those in CEL, implying that as professionally trained educational practitioners, Chinese EMI lecturers are, at least, aware of lecture genres as well as native English lecturers. On the other hand, Chinese lecturers seem to rely heavily on certain moves/steps in classroom lecturing while neglecting others.

Then, the researcher adopted intuition-oriented criteria, including *grammatical irregularity and/or semantic opacity* and *Morpheme Equivalent Unit* rule, to identify formulaic sequences. The log-likelihood comparison (Liang, Li & Xu, 2010) reveals significantly fewer types (log-likelihood=-25.83, p=0.00<0.05) and tokens (log-likelihood=-7.02, p=0.01<0.05) of formulaic sequences used in CCL lectures than in CEL lectures. Meanwhile, Chinese EMI lecturers also seem to rely heavily on very

limited types of formulaic sequences.

Three student interviewees and one lecturer interviewee in the pilot study explicitly stated that non-standard use of formulaic sequences such as *we can got*, *we can saw*, *at the beginning* (instead of *in the beginning* or *at the beginning of*) would not at all cause any communication breakdown.

3.6.2 Revisions after the pilot study

In order to conduct the main study in a more principled way, several revisions were made after the pilot study.

The first change is to rename ***Topic Network Building Phase*** into ***Theme Network Building Phase*** since 'theme' could be more appropriate to cover all topics involved in lectures. Then, some moves were discarded while new moves were added in the framework (Figure 3.3).

Topic Network Building Phase (Initial)
M4: Setting Up Topic Framework
M5: Putting Topic In Context
M6: Topic Knowledge Checking

→

Theme Network Building Phase (Revised)
M4: Introducing The Topic
M5: Elaborating The Topic
M6: Building Theme Network
M7: Making Aside
M8: House Keeping
M9: Checking Comprehension & Consolidating Learning
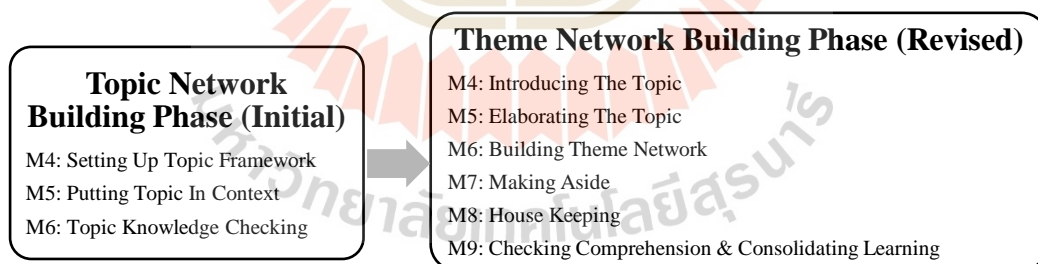
Figure 3.3 Revision of moves

The next revision concerns guiding questions for interviews, which were radically revised and redesigned on the basis of initial genre analysis of lecture transcripts according to the feedback from committee members.

Lastly, the influence of Covid-19 renders it difficult to contact students.

Therefore, the main study did not involve interview with students, instead, it only interviewed Chinese EMI lecturers as to their conceptualization on rhetorical move structure as well as their attitudes towards formulaic language use, non-standard FS use in particular.

3.7 Summary

This chapter gives a detailed description of the methodology employed for the present research. It begins with a presentation of the overall design of the research, specifying the research questions. Then it details the principles and procedures in the creation of the specialized corpora of CCL and CEL. After that, it justifies the top-down approach to move analysis framework for the present research. Then, the identification and structural categories of formulaic sequences as well as their relation with moves/steps were elaborated. Finally, the chapter ends with a report of the pilot study.

CHAPTER 4

RESULTS AND DISCUSSION I: COMPARISON OF

RHETORICAL MOVE STRUCTURE IN CCL AND CEL

This chapter presents the results and discussion of move analysis conducted on EMI lectures in CCL and CEL. First, it provides an overview of rhetorical move structure identified in EMI lectures of CCL and CEL. Then, the major part of this chapter compares all the moves/steps in EMI lectures of the two corpora. The following section examines the move structure of EMI lectures from an ELF perspective. The chapter ends with a summary of the results and findings of rhetorical move structure comparisons.

4.1 Overview of rhetorical moves in CCL and CEL

The comparison of rhetorical move structure in EMI lectures of CCL and CEL are expected to answer research question 1) proposed in Chapter 1, which is presented again as follows:

1) What are the similarities and differences in rhetorical move structure of university EMI lectures given by native and non-native English lecturers?

To address research question 1), a move analysis framework on the basis of Lee

(2016) was developed for the analysis of EMI lectures in CCL and CEL. However, it must be noted that linguistic features and/or devices such as personal pronouns, metadiscourse and questions are generally not dealt with in this part, but instead will be presented in combination with the use of formulaic language in Chapter 5 for reasons suggested in Section 2.2.6.

Table 4.1 Overall moves/steps statistics

Moves	Distribution (percentage) [a]		Frequency [b]		LL critical value G2 [c]
	CCL	CEL	CCL	CEL	
M1 Getting Started	12 (100%) [#]	5 (42%)	12	5	+ 8.91[**]
M2 Warming Up	7 (58%) [#]	12 (100%) [#]	13	29	- 0.31
M3 Setting Up Lecture Agenda	12 (100%) [#]	12 (100%) [#]	22	18	+ 6.65[**]
M4 Introducing The Topic	12 (100%) [#]	12 (100%) [#]	112	162	+ 4.07[*]
M5 Elaborating The Topic	12 (100%) [#]	12 (100%) [#]	420	664	+ 6.59[**]
M6 Building Theme Network	11(92%) [#]	12 (100%) [#]	55	156	- 7.76[**]
M7 Making Aside	2 (17%)	9(75%) [#]	4	19	- 3.51
M8 Housekeeping	3 (25%)	10(83%) [#]	3	40	- 19.00[****]
M9 Checking Comprehension & Consolidating Learning	5(42%)	6 (50%) [#]	8	31	- 3.93[*]
M10 Wrapping Up Lecture	7 (58%) [#]	2 (17%)	8	2	+ 8.51[**]
M11 Cooling Down	8 (67%) [#]	10 (83)% [#]	13	19	+ 0.44
M12 Ending Lecture	12 (100%) [#]	12 (100%) [#]	15	16	+ 2.34
Total			685	1161	+ 3.58

Note. [a] Distribution refers to the number of lectures where certain moves/steps are present. [b] Frequency refers to the number of moves/steps that are identified. [c] The log-likelihood (LL) critical value G2 itself is always a positive number. However, indicators '+' and '-' are inserted to show overuse and underuse of corpus 1 (in this case CCL) relative to corpus 2 (in this case CEL). [*]$p < 0.05$, critical value G2= 3.84; [**]$p < 0.01$, critical value G2 = 6.63; [***]$p < 0.001$, critical value G2 = 10.83; [****]$p < 0.0001$, critical value G2 = 15.13. [#] indicates conventional status of moves/steps.

Table 4.1 shows the overall distribution and frequency of moves/steps identified in EMI lectures of CCL and CEL.

The table reveals that all 12 moves appear in both CCL and CEL. In terms of distribution, nine moves in CCL and ten moves in CEL reach the conventionality level of 50% as defined in the present study. Seven of them share the same conventional move status. Specifically, *M2 Warming Up* and *M3 Setting Up Lecture Agenda* in the Opening Phase, *M4 Introducing The Topic*, *M5 Elaborating The Topic* and *M6 Building Theme Network* in the Theme Network Building Phase, as well as *M11 Cooling Down* and *M12 Ending Lecture* in the Closing Phase, are conventional in both corpora. The overall move distribution and conventionality seem to suggest that CCL and CEL share similar rhetorical move structure.

With regard to actual occurrences, a total of 685 and 1161 moves/steps were identified in CCL and CEL, respectively. Log likely-hood comparison shows no significant difference (G2= +3.58, p>0.05) in terms of the frequency of moves/steps in EMI lectures given by native and non-native English lecturers. In fact, the critical value G2 +3.58 indicates slightly more moves/steps used in EMI lectures given by Chinese lecturers than their native English counterparts, indicating that they are aware of the rhetorical move structure of lectures, at least, as well as their native English counterparts. In fact, all four interviewees admitted that the generic structure of their EMI lectures are quite similar to that of their Chinese lectures. In particular, Interviewee A articulated that both his Chinese lectures and EMI lectures were strictly guided by similar teaching plans and course syllabus, and Interviewee B affirmed that mathematics courses share similar patterns whether delivered in English or Chinese.

Nevertheless, differences were found in individual moves/steps between EMI lectures in CCL and CEL, with the major discrepancy centering upon the Opening Phase and the Theme Network Building Phase.

In terms of conventionality, *M1 Getting Started* and *M10 Wrapping Up Lecture* are conventional in CCL, but optional in CEL; on the contrary, *M7 Making Aside*, *M8 Housekeeping* and *M9 Checking Comprehension & Consolidating Learning* are optional in CCL, but conventional in CEL.

In terms of frequency of moves/steps, there are significantly more *M1 Getting Started*, *M3 Setting Up Lecture Agenda*, *M4 Introducing The Topic, M5 Elaborating The Topic* and *M10 Wrapping Up Lecture* in CCL than in CEL, whereas there are significantly more *M6 Building Theme Network*, *M8 Housekeeping* and *M9 Checking Comprehension & Consolidating Learning* in CEL than in CCL.

The following sections present thorough comparisons of all the moves/steps in individual phases of EMI lectures in CCL and CEL.

4.2 Rhetorical moves in the Opening Phase of EMI lectures in CCL and CEL

The Opening Phase serves to orient students to the lecture by signaling the official start, informing students of lecture-related issues, and setting up lecture agenda. Generally, the Opening Phase of EMI lectures are not dissimilar to the Opening Phase of EAP lectures identified by Lee (2016), though with different moves/steps. Three

moves were found in the Opening Phase, viz., *M1 Getting Started*, *M2 Warming Up* and *M3 Setting Up Lecture Agenda*.

Table 4.2 Moves/steps in the Opening Phase of EMI lectures

Moves	Distribution [a]		Frequency [b]		LL critical value G2 [c]
	CCL	CEL	CCL	CEL	
M1 Getting Started	12 (100%) [#]	5 (42%)	12	5	+ 8.91[**]
M2 Warming Up	7 (58%) [#]	12 (100%) [#]	13	29	- 0.31
M2S1 Leading in	6 (50%) [#]	0 (0%)	8	0	+ 16.80[****]
M2S2 Recalling previous lecture(s)	3 (25%)	9 (75%) #	3	11	- 1.23
M2S3 Looking ahead	0 (0%)	3 (25%)	0	4	- 3.45
M2S4 Housekeeping	1 (8%)	8 (67%) [#]	2	14	- 4.20[*]
M3 Setting Up Lecture Agenda	12 (100%) [#]	12 (100%) [#]	22	18	+ 6.65[**]
M3S1 Announcing/Clarifying lecture theme	11 (92%) [#]	8 (67%) [#]	12	13	+ 1.78
M3S2 Providing lecture scope	9 (75%) [#]	4 (33%)	10	5	+ 6.21[*]

Note. [a] Distribution refers to the number of lectures where certain moves/steps are present. [b] Frequency refers to the number of moves/steps that are identified. [c] The log-likelihood (LL) critical value G2 itself is always a positive number. However, indicators '+' and '-' are inserted to show overuse and underuse of corpus 1 (in this case CCL) relative to corpus 2 (in this case CEL). [*]$p < 0.05$, critical value G2= 3.84; [**]$p < 0.01$, critical value G2 = 6.63; [***]$p < 0.001$, critical value G2 = 10.83; [****]$p < 0.0001$, critical value G2 = 15.13. [#] indicates conventional status of moves/steps.

Table 4.2 presents distribution and frequency data of each move/step in the Opening Phase.

All moves/steps (except *M2S3* in CCL and *M2S1* CEL) in this phase appear in both corpora. According to the conventionality criterion stipulated in the present study, all, except *M1 Getting Started* in CEL, are conventional. This seems to indicate similar rhetorical move structure in the Opening Phase of the two corpora.

M1 Getting Started signals the official start of a lecture, and it is often realized by a combination of discourse markers, greetings, directives, and rhetorical questions, as shown in excerpt 4-5). These expressions are common linguistic resources for teachers to start lectures and orient students to the lecture (Biber & Conrad, 2009; Wong Fillmore, 1985).

4) *S1: **OK**, let's get it started. Good afternoon, everyone!* (CCL08)

5) *S1: okay why don't we get started,* (CEL12)

There are significantly more ***M1*** in CCL than in CEL (G2= + 8.91, p<0.01). ***M1*** was observed in all CCL lectures, which shows the conventionality of explicit signals to start lectures in the Chinese tertiary education context. Therefore, it is a conventional move in CCL. However, ***M1*** only occurred in five CEL lectures, thus could be an optional move in CEL. Among these five lectures that include ***M1***, two (CEL03 and CEL10) of them didn't start with ***M1***, but instead lecturers in these two lectures start their class with ***M2S4 Housekeeping***. This could be an indication of the non-sequential order of moves/steps in academic lectures. In addition, among the other seven lectures that do not include ***M1***, three of them start with ***M2S2 Recalling previous lecture(s)*** (CEL01, CEL02 and CEL06), three with ***M2S4 Housekeeping*** (CEL05, CEL08 and CEL09), and one with ***M3S1 Announcing/Clarifying lecture theme*** (CEL11). This may suggest that an official start of a lecture through ***M1*** is only optional for native English lecturers, depending on individual lecturers in various situations. This was also evidenced in Thompson (1994), where no similar Function/Sub-Function (move/step)

was identified in her analysis of lecture introductions. However, Lee (2011, 2016) identified *M1: Getting Started* as conventional in his study of EAP lectures. A reasonable speculation could be that language lecturers tend to verbally announce the start of lectures, but discipline lecturers do not seem to care about such explicit announcements. Chinese EMI lecturers seem to be prone to practices of native English language lecturers in *Move 1*.

M2 Warming Up serves to lead students to the present lecture through presenting lead-in information, recalling previous lecture, looking ahead to future lecture, and highlighting course-related issues before engaging in more substantive parts of the lecture. Four steps were identified in *M2*, including *M2S1 Leading in*, *M2S2 Recalling previous lecture(s)*, *M2S3 Looking ahead* and *M2S4 Housekeeping*. This differs considerably from the three steps of *M2: Warming up* of EAP lectures identified by Lee (2011, 2016), which include *S1: Housekeeping*, *S2: Looking ahead*, and *S3: Making a digression*. In addition, Thompson (1994) didn't include *M2 Warming Up* in her analysis of lecture introductions. The differences are illustrated in the discussion of individual steps below.

One conventional step and two optional steps were identified in CCL, whereas two conventional steps and one optional step were found in CEL. Generally, no significant difference (G2= -0.31, p>0.05) was found in *M2* between CCL and CEL, except for the substantially different preferences in the use of *M2S1 Leading in* by lecturers of the two corpora.

M2S1 Leading in is conventional in CCL, but it does not occur in CEL at all. There are significantly more *M2S1* (G2= +16.80, p<0.01) in CCL than in CEL. *M2S1* is frequently used by Chinese lecturers to help prepare students for the lecture, usually with examples/issues directly or indirectly related to the course content. It contextualizes students in an imagined scenario and it often uses a rhetorical question to connect the imagined situation with the lecture theme (see excerpt 6)).

6) *You know, when you normally walk in a famous area, and you see some nice, you know, landscapes, normally you will be attracted by different kind of rocks. They have different colors, different shapes, and different compositions if you can recognize. And so what is rocks?* (CCL02)

M2S1 Leading in is common practice in Chinese lectures, which might have been transferred by Chinese lecturers when they had to teach EMI courses. This could be evidenced in the interview data where all interviewees admitted such transfer. Unexpectedly, no *M2S1* was found in the CEL lectures. However, Lee (2011, 2016) identified an optional step *M2S3: Making a digression* which may allow teachers to discuss real or imagined issues less germane to course content. This deceptively-semblable step is very different from the step *M2S1 Leading in* in the present study. Though both involve discrete situations that may or may not be real, *M2S1 Leading in* in the present study is to make natural pavements for and connections to the lecture theme, while *M2S3: Making a digression* in Lee (2011, 2016) aims at sustaining positive teacher–student relationship.

In **M2S2 Recalling previous lecture(s)**, lecturers present contents of previous lecture(s) to refresh students' memory so as to prepare them for the present lecture (see excerpt 7-8)).

7) Let's *see, last class in this section chapter two, we talked about conventional artificial lift methods, such as like sucker rod pumps, ESP, Gas lift and PCP.* (CCL06)

8) we've *looked so far, at the design of fishes, the way they feed, and those kinds of issues, and we started last time thinking about food, and feeding, and the physiological, capabilities of fishes particularly in terms of the way they process energy.* (CEL08)

M2S2 is optional in CCL, but conventional in CEL. No significant difference (G2= -1.23, p>0.05) was found in the use of this step between lectures in CCL and CEL. The difference in conventionality might be related to the course syllabus and delivery of the lectures in CCL. The twelve lectures selected from the four courses (*An Introduction to Earth Sciences*, *Recent Advances in Oil and Gas Production Engineering*, *Advanced Mathematics I* and *Oil Drilling Engineering*) were all co-taught by different lecturers, therefore may undermine the course continuity to certain degrees, leading to relatively lower frequency of **M2S2 Recalling previous lecture(s)** in CCL. In addition, Interviewee C revealed that her lectures were part of the comprehensive course, which cover earth science, petrology, geological history, and paleontology. Each part could be an independent course, but due to the introductory nature of the course, was condensed

in one lecture course. Therefore, there was little continuity shown between lectures in her course.

It should be mentioned, neither Lee (2011, 2016) nor Thompson (1994) identified the step *M2S2 Recalling previous lecture(s)* in their studies. On the one hand, it might be related to the pedagogical orientations of lectures, since the research objects in Lee (2011, 2016) were EAP lectures whose primary aim is language teaching and learning, which does not necessitate a continuity in the knowledge content. On the other hand, though Thompson (1994) examined academic lectures from a range of discipline areas (*applied linguistics*, *engineering*, and *medicine*), nine of the eighteen lectures were 'one-offs', which does not require continuity at all.

M2S3 Looking ahead functions to inform students of the upcoming lecture(s) (see excerpt 9)).

9) *this week and next you're gonna see some extensions where we look at alternative I-O point locations in the rack.* (CEL10)

No *M2S3* was found in CCL, and it only occurred in three lectures (25%) in CEL, thereby is optional in CEL. No significant difference (G2= -3.45, p>0.05) was found in the use of *M2S3 Looking ahead* between the two corpora. Even in CEL, this step appeared only in three lectures (25%), which is similar to the 29.2% distribution of EAP lectures in Lee (2011, 2016). All these seem to indicate that the step *M2S3 Looking ahead* is quite optional, whether for native or non-native English lecturers, and whether in subject lectures or EAP lectures.

M2S4 Housekeeping is used to make class management announcements, and offer reminders (see excerpt 10-12)).

10) *I'll try to make you understand it. If you have any question, any question, please ask me, ok?* (CCL06)

11) *what i wanted to say to you this morning [0.3] is that i think w-, [0.5] today's session is going to be [0.3] the last [0.3] of the lecture sessions [0.4] for the course [0.6] or if it isn't there'll only be about ten or fifteen minutes next week [0.4]* (CEL02)

12) *alright as you see we've got a slightly unusual, event today we're being taped, and uh, basically we'll just pretend they're not here which is_ was their request, uh before we get going with the selection sort again are there any questions about anything...?* (CEL12)

M2S4 appeared only in one CCL lecture, thus is optional in CCL; nevertheless, it was used in eight CEL lectures, thereby is conventional in CEL. Significantly less (G2= -4.20, p>0.05) **M2S4** was found in lectures of CCL than lectures of CEL. This seems to suggest that native English lecturers may have the inclination to include such a step in their lectures so as to get across their intentions, as in excerpt 11), and to sustain positive and close relationship with students, as in except 12). On the other hand, Chinese lecturers do not always seem to favor this step to do classroom management. Interviewee A certainly was confident of his lecture, and skipped the question. Interviewee B mentioned that classroom management was closely related to the time

allocation of the class, and Interviewee D said classroom management thing was partly done by the administrative staff. Only Interviewee C talked of randomly roll-calling and homework checking.

It is worth noting that Lee (2011, 2016) also identified the ***Housekeeping*** step as conventional. It seems that native English lecturers, be they discipline or language teachers, have the tendency to use ***M2S4 Housekeeping*** to help inform students of lecturers' intentions and maintain close teacher-student relationship. On the contrary, Chinese lecturers do not seem to favor close relationship with students due to the influence of traditional culture which advocates respect for and authority of teachers.

M3 Setting Up Lecture Agenda is used to brief contents to be covered in the present lecture. Two steps were identified, viz., ***M3S1 Announcing/Clarifying lecture theme*** and ***M3S2 Providing lecture scope***.

In the similar move structure which Thompson (1994) termed as ***Function: Set up the Lecture Framework***, she identified four Sub-Functions: 1) *Announce topic (15)*, 2) *Indicate scope (12)*, 3) *Outline structure (11)*, and 4) *Present aims (9)*. These Sub-functions are quite frequent, with the number in the brackets indicating frequency of these steps in the 18 lectures in her study. However, Sub-Function 1) and 4), as defined in her study, are always intertwined, therefore were subsumed under ***M3S1 Announcing/clarifying lecture theme*** in the present study. Analogously, Sub-Function 2) and 3) were merged as ***M3S2 Providing lecture scope***.

M3S1 Announcing/Clarifying lecture theme is used to explicitly introduce the

lecture theme to students (see excerpt 13-14)).

13) *Today we will talk about the Chapter Four, igneous rocks.* (CCL02)

14) *okay so we [1.3] move on [1.9] with [1.7] the reinforced materials [1.4]* (CEL03)

No significant difference (G2= +1.78, p>0.05) was found in the use of **M3S1** between lectures in CCL and CEL, and it is conventional in both corpora. Similar finding was also found by Thompson (1994), where this step occurred, among the total 18 lectures of her study, as Sub-Function 1) in 15 (83%) lectures and as Sub-Function 4) in 9 (50%) lectures. However, Lee (2011, 2016) didn't have the topic announcement move/step due perhaps to the language-orientations of the EAP lectures.

M3S2 Providing lecture scope enumerates contents to be covered in the lecture, as can be shown in excerpt 15-16).

15) *Em for this chapters, we will learn nine parts, but today this class we just learn the former fourth parts: geological resources, nonmetallic mineral resources, metals and ore, and how ore forms.* (CCL03)

16) *okay. well what i'd like to do first... is review the algorithm, we did go through it yesterday but i think this is uh tough enough to understand that we probably oughta go through it, uh, quicker this time but go through it again, and then when we look at the code, i think it'll be easier to understand if we've just looked at these, uh graphics again to see how the sort works.* (CEL12)

There are significantly more (G2= +6.21, p<0.05) use of **M3S2** in CCL than in

CEL. It is conventional in CCL, but optional in CEL. Thompson (1994) had consistent findings concerning the frequency of this step, where it appeared, among the 18 lectures investigated, as Sub-Function 2) in 12 (67%) lectures and as Sub-Function 4) in 11 (61%) lectures. Nevertheless, the functionally comparable move *M3: Setting Up Lesson Agenda* in Lee (2011, 2016) is only optional, occurring in 29.2% (7/24) of EAP lectures in his study.

It seems to suggest that Chinese discipline lecturers favor the inclusion of explicit proclamation of the lecture scope so as to lay out the structure to students. All four interviewees confirmed the conventional use of this step to inform students of lecture scope in the interviews. Nevertheless, findings about the native lecturers seem to be split in different studies. While earlier study by Thompson (1994) showed the preference of *M3S2* as a conventional step by the native discipline lecturers, native language lecturers in Lee (2011, 2016) and native discipline lecturers of MICASE and BASE in the present study tend to use this step optionally.

4.3 Rhetorical moves in the Theme Network Building Phase of EMI lectures in CCL and CEL

As the main part of academic lectures, Theme Network Building Phase consists of a series of recursive and non-sequential moves/steps that teachers (with students) employ to help build a knowledge network which facilitates learning of knowledge content. Theme is the core of a lecture, upon which the whole teaching was centered,

whereas topics are the main points that are essential for building the corresponding theme. Simply put, a lecture theme is generally supported and forged by a number of topics, relevant to various degrees. Usually, a typical lecture deals with one theme, which was often indicated by the lecture title. In essence, themes and topics are two essential elements of academic lectures.

There are six moves identified in the Theme Network Building Phase, viz., **M4 Introducing The Topic, M5 Elaborating The Topic, M6 Building Theme Network, M7 Making Aside, M8 Housekeeping,** and **M9 Checking Comprehension & Consolidating Learning**.

Table 4.3 Moves/steps in the Theme Network Building Phase

Moves	Distribution [a]		Frequency [b]		LL critical value G2 [c]
	CCL	CEL	CCL	CEL	
M4 Introducing The Topic	12 (100%) [#]	12 (100%) [#]	112	162	+ 4.07[*]
M5 Elaborating The Topic	12 (100%) [#]	12 (100%) [#]	420	664	+ 6.59[**]
M5S1 Explaining terms	7 (58%) [#]	5 (42%)	16	12	+ 5.69[*]
M5S2 Highlighting importance	1 (8%)	4 (33%)	1	6	- 1.53
M5S3a Presenting background knowledge	11 (92%) [#]	12 (100%) [#]	51	64	+ 4.27[*]
M5S3b Contextualizing real-word experience	4 (33%)	9 (75%) [#]	4	26	- 7.24[**]
M5S3c Setting up premise	1 (8%)	6 (50%) [#]	3	11	- 1.23
M5S4 Expounding rationale	12 (100%) [#]	11 (92%) [#]	129	89	+52.71[****]
M5S5 Demonstrating the topic	12 (100%) [#]	12 (100%) [#]	141	231	+ 1.36
M5S6 Providing caveats	3 (25%)	10 (83%) [#]	9	43	- 8.03[**]
M5S7 Making comments	7 (58%) [#]	12 (100%) #	19	74	- 9.47[**]
M5S8 Summing up the topic	5 (42%)	9 (75%) [#]	9	14	+ 0.17
M5S9 Initiating co-building lecture	8 (67%) [#]	12 (100%) [#]	36	58	+ 0.44
M5S10 Pinning down and/or clarifying the topic	2 (17%)	11 (92%) [#]	2	36	- 19.54[****]

Table 4.3 Moves/steps in the Theme Network Building Phase (Cont.)

Moves	Distribution [a]		Frequency [b]		LL critical value G2 [c]
	CCL	CEL	CCL	CEL	
M6 Building Theme Network	11 (92%) [#]	12 (100%) [#]	55	156	- 7.76[**]
M6S1 Referring to previous lecture/other Source	8 (67%) [#]	12 (100%) [#]	13	72	- 16.60[****]
M6S2 Specifying subsequent/future content	4 (33%)	8 (67%) [#]	8	30	- 3.53
M6S3 Connecting/ Comparing topics	11 (92%) [#]	11 (92%) [#]	34	54	+ 0.50
M7 Making Aside	2 (17%)	9 (75%) [#]	4	19	- 3.51
M8 Housekeeping	3 (25%)	10 (83%) [#]	3	40	- 19.00[****]
M9 Checking Comprehension & Consolidating Learning	5 (42%)	6 (50%) [#]	8	31	- 3.93[*]

Note. [a] Distribution refers to the number of lectures where certain moves/steps are present. [b] Frequency refers to the number of moves/steps that are identified. [c] The log-likelihood (LL) critical value G2 itself is always a positive number. However, indicators '+' and '-' are inserted to show overuse and underuse of corpus 1 (in this case CCL) relative to corpus 2 (in this case CEL). [*]$p < 0.05$, critical value G2= 3.84; [**]$p < 0.01$, critical value G2 = 6.63; [***]$p < 0.001$, critical value G2 = 10.83; [****]$p < 0.0001$, critical value G2 = 15.13. [#] indicates conventional status of moves/steps.

Table 4.3 shows the distribution and frequency data of each move/step in the Theme Network Building phase.

All moves/steps in this phase appear in both CCL and CEL. However, some differences in conventionality were found. *M4*, *M5* and *M6* are conventional, but *M7*, *M8* and *M9* are optional in CCL. Meanwhile, all moves are conventional in CEL.

In ***M4 Introducing The Topic***, lecturers introduce topics germane to the lecture theme. It is usually realized by explicit commissives 17) or rhetorical questions 18) of the lecturer.

17) *And we can, we will talk about, more about the San Andreas Fault*. (CCL01)

18) *what about prognostic indicators [1.2] well [3.2] one of the [0.6] best*

prognostic indicators still [1.2] is the level of your C-D-four count [3.7]

(CEL06)

M4 occurs in all lectures in both CCL and CEL, and it is conventional for both

native and non-native English lecturers. There are significant more use of *M4* (G2=

+4.07, p<0.05) by lectures of CCL than those of CEL. A likely explanation could be

that Chinese EMI lecturers made less thorough elaboration on individual topics, instead

they may favor providing students with more related topics in certain areas. With regard

to the use of *M4*, the four Chinese EMI lecturers are quite split in the interviews.

Interviewee A and Interviewee B replied that they tend to introduce more related topics

in lecturing, depending on the course level and student level. However, Interviewee C

and Interviewee D claimed they preferred probing deeper into certain topics to

introducing more related areas of topics. In sum, Chinese EMI lecturers' use of *M4*

could be jointly influenced by course level, student level, as well as their personal

lecturing style.

Lee (2011, 2016) didn't include this step in his study since the research objects are

EAP lectures which is, to a great extent, language-oriented. Though Thompson (1994)

identified the ***Sub-Function: Announce Topic***, she didn't distinguish between the

lecture theme and topics supporting the theme, which was implied in the elaboration of

the ***Sub-Function***, where "the lecturer gives the title or topic of the lecture" (p. 176).

M5 Elaborating the topic is concerned with various strategies lecturers may

adopt to get across the topics introduced. As the most important part of academic lectures, *M5* is realized through ten different steps, among which *M5S3* is realized by three sub-steps. *M5* appear in all lectures examined, thus are conventional in both CCL and CEL. In addition, log-likelihood comparison shows significantly more (G2= +6.59, p<0.05) use of this move in CCL than in CEL. Nevertheless, the constituent steps and sub-steps exhibit different levels of conventionality and frequency between native and non-native English lecturers in CCL and CEL.

    M5S1 Explaining terms is used to give detailed explanation of technical terms involved in the topic (see excerpt 19-20)).

19) *Ok, **geological resource** can be defined as concentration of naturally occurring solid, liquid, or gas material in or on the Earth's crust in such forms and amount that economic extraction of a commodity from the concentration is currently or potentially feasible.* (CCL03)

20) *and the **G value** [0.6] which is a measure of the product yield or indeed the react-, the [0.2] destruction of the reactant 'cause er that would be then the negative G value but [0.5] they normally normally you look at products to measure positive G values [1.0] they are the number of molecules [0.2] formed [0.3] or destroyed [0.5] for every hundred electron volts [2.7]* (CEL01)

    M5S1 appears in 58% (conventional) of CCL lectures, but in 42% (optional) of CEL lectures. The critical value G2 indicates significantly more (G2= 5.69, p<0.05) use of this step by Chinese lecturers than their native counterparts. It could be seen in

excerpt 19) that the Chinese lecturer gave a relatively complete and cohesive definition of the technical term '*geological resource*', which could be an indication of reliance on textbook definitions or prepared power-point slides. This was verified by the teaching video where they often resort to long periods of definition reading. Further evidence could be found from the interviews where all four interviewees admitted referring to authoritative definitions from textbooks and published articles while explaining technical terms. On the other hand, in excerpt 20) the native English lecturer's definition of the term '*G value*' showed traces of self-repair (e.g., *the react-, the [0.2] destruction of*), use of filler (e.g., *'cause er*), repetition (e.g., *they normally normally*) as well as incoherence (e.g., *they normally normally you look at products*). These could be signs of on-line processing of information, showing relatively little reliance on prepared written material.

M5S2 Highlighting importance highlights importance of the topic. It is often symbolized by the use of adjectives denoting importance, such as 'important', 'critical', 'key', or comparative/superlative adjective forms (see excerpt 21-22)).

21) *But to understand the mechanism is so **important**, so I'll talk about why we has to understand the mechanism… So, so that's why the model, they make mechanism is so important for gas production engineers.* (CCL06)

22) *water is probably **the most important** system looked at [0.5] mainly because as i said [0.4] we are [0.6] we consist largely of water [0.8] so there's huge*

interest [0.3] there must must be over a thousand papers on the radiation

chemistry of water and probably half a dozen books as well [1.0] (CEL01)

No significant difference (G2= -1.53, p>0.05) was found in the use of **M5S2**

between lectures in CCL and CEL. This step was used with low frequency, occurring

only once in CCL lectures and six times in four CEL lectures respectively. It is optional

in both corpora. This echoes the findings of Lee (2009), which reported only one

occurrence of **Move 3 Step 1: Showing the importance of the topic** in small class

lectures and two occurrences in large class lectures in his study, thus were optional.

However, a similar step in Thompson (1994) **Sub-Function: Show importance/**

relevance of topic, which was used to highlight the particularly interesting, central, or

widespread features of the lecture topic, appeared with relatively high frequency in her

study, occurring in 11 out of the 18 lectures. This could be, to a more or lesser extent,

related to the different orientations of language classes and discipline lectures. In this

aspect, Chinese EMI lecturers seem to resemble more of native English language

lecturers than discipline teachers.

M5S3 revolves around the background of the topic under discussion. Three

different sub-steps were identified in the realization of this step, viz., **M5S3a**

Presenting background knowledge, **M5S3b Contextualizing real-word experience**,

M5S3c Setting up premise. These sub-steps were used differently by lecturers. Among

the three sub-steps, **M5S3a** was most frequently used by lecturers in both CCL and CEL,

followed sequentially in terms of frequency by **M5S3b** and **M5S3c** in the two corpora.

As the most frequently used sub-step, ***M5S3a Presenting background knowledge*** is used by lecturers to directly present background knowledge of the topic. It is often realized through a series of factual statements which are considered essential for the understanding of the topic (See except 23-24)).

23) *Yeah, em, at that time the Earth was controlled by all kind of dinosaurs, in the ocean, in the land, in the air, atmosphere, all kinds of, yeah atmos- er, during this time, during this time, we will talk about this, the earliest bird came into the world. We called that archaeopteryx, archaeopteryx, em the first archaeopteryx fossil was discovered in Germany in nineteen fifty-two. The fossil is really much small, almost, you can find the picture in the internet. In Chinese, we called it Shizuniao, but in English we call it archaeopteryx. Er, it's much more like that, er this fossil was kept in the famous museum in Germany. Sometimes when you see the picture of the fossil, most of us guess it was reptile, em not like the bird, like our like today, but it was the earliest bird.* (CCL01)

24) *once H-I-V once AIDS became clearly an important [0.3] progressive ecidem-, e-, [0.3] epidemic [1.3] the World Health Organization set up [0.7] a task force [0.4] to [0.5] measure the impact to measure the epidemic and its impact globally [0.5] and this task force now produces a [0.6] a global report [0.2] annually in December [0.9] and so what i'm going to show you [0.2] are the figures from December nineteen-ninety-seven [0.5] and in a month's or so's time [0.3] there will be figures coming out for December nineteen-ninety-eight*

[17.1] right this gives you an indication [0.7] of the W-H-O's estimate [0.5] of the number of adults and children [0.8] estimated to be living with H-I-V infection [0.8] at the end of nineteen-ninety-seven [2.0] (CEL06)

In excerpt 23), in order to elaborate the geological features of Jurassic period, the lecturer in CCL01 discussed the omniscient presence of dinosaurs as well as the arrival of archaeopteryx. And in excerpt 24), the lecturer introduced the efforts of World Health Organization and the global report of December nineteen-ninety-seven to get students to understand the global situation of AIDS.

M5S3a appears in 92% (11/12) and 100% (12/12) of lectures in CCL and CEL respectively, thus is conventional in both corpora. Nevertheless, log-likelihood comparison shows significantly more (G2=+ 4.27, p<0.05) use of this sub-step by native Chinese lecturers than native English lecturers. This might be a result of over-reliance of Chinese lecturers on this single type of background-related sub-step. *M5S3a* account for 88% (51/58) of all background-related sub-steps in CCL, while it only takes up 63% (64/101) of background sub-steps in CEL.

M5S3b Contextualizing real-word experience involves referring to the real-word experience of students as the background to help understand the topic introduced (See except 25-26)).

25) *Er you know, people, people are closed relationship with the natural resources for every day, daily life. Yeah? Yeah. Today, with silicon chip that operate our computer, just like this. Every computer has this, yeah. And the gasoline that*

power our car, all of this, the d- product from earth's resources. But we use the

geological resources very long time. Pre-history's ancient people can use this,

em, this flint to make a flint knife, or use obsidian scraper, yeah. (CCL03)

26) *now if someone [0.7] if someone when they're mowing the lawn on the ga-, on*

the [0.3] on the [0.4] on the site manages to bash their rain gauge with a [0.6]

er [0.2] lawnmower and i gather that's not er uncommon [0.5] then you you

end up with your [laugh] [0.4] with your rain gauge being something less than

[0.5] less than circular [0.5] so there's there's [0.2] there's various different er

[1.4] different [0.8] er [0.7] sources of error [6.2] (CEL04)

In excerpt 25), the Chinese lecturer in CCL03 used such common things as 'silicon chip' and 'gasoline' in daily life to contextualize the concept of geologic resource in the real world, which may help students' understanding of the relatively abstract concept that is rare in life. Similarly, the concept of 'measurement accuracy of rainfall gauge' may not be that common, the native English lecturer in except 26) of CEL04 related the daily experience of lawn mowing to factors that may affect the measurement accuracy, which could be much easier for students.

M5S3b appears in 33% (4/12) and 75% (9/12) of lectures in CCL and CEL respectively, thus is optional in CCL, but conventional in CEL. Significant difference (G2= -7.24, p<0.01) was reported in the use of this sub-step, with Chinese lecturers of CCL using significantly less *M5S3b* than CEL lecturers. This might be seen as the difference in the capacity and/or preference of lecturers to relate the abstract to the

concrete, and to relate book knowledge to life experience.

M5S3c Setting up premise is to present the known facts or conditions as the foundation for understanding a certain topic. It is often exhibited in lectures that favor problem-solution patterns of instruction, for example *Mathematics* as in excerpt 27) and *Engineering* as in excerpt 28).

27) *Now, er, we know that a particle, a particle have the rectilinear motion and we use s equal f of t to denote the position function, you know position function? Now look at this figure, look at this figure. This is, er, origin. This is s axis. And if we know time t0, in this time t0, the position function is f(t0), and if we give the increment, now look at this, in time t0 plus h, so we have correspond f (t0 plus h) to denote the position function at this time, ok.* (CCL11)

28) *we know the properties of the matrix [1.0] in terms of [1.2] that's Young's modulus [1.0] Poisson's ratio [1.0] sheer modulus [1.1] similiarly [0.4] we do know the properties of the fibres [1.0] again Young's modulus [0.9] Poisson's ratio [1.1] and sheer modulus [1.8] and so long as we know how many fibres we've actually put [0.2] in the system [0.9] okay [2.8] so we've got fibre content [1.9] okay we can try to see how we [0.3] work on that [1.0]* (CEL03)

In excerpt 27), in order to solve the 'position function' of a particle, the Chinese lecturer presented all known conditions, such as 'a particle have rectilinear motion', the axis, the time, and increment. In excerpt 28), the native English lecturer listed all the known factors, such as matrix property, Young's modulus, Poisson's ratio, sheer

modulus and fiber content, so as to illustrate the mechanism to embed fibers in the matrix.

M5S3c appears in 8% (1/12) and 50% (6/12) of lectures in CCL and CEL respectively, thus is optional in CCL, but conventional in CEL. No significant difference (G2= -1.23, p>0.05) was found in the use of this sub-step between the two corpora. Generally, this sub-step was not used frequently in either CCL or CEL.

The use of three background-related sub-steps seem to denote a tentative conclusion that native Chinese lecturers tend to rely on particular sub-step, in this case *M5S3a*, to present background, while the native English lecturers seem to have wider choices in this regard. All four interviewees verified their inclination to use *M5S3a* as the sole strategy to present background information. Nonetheless, all except interviewee D claimed their preference of *M5S3b*, which contradicts text analysis results. It is likely that these interviewee lecturers may intend to use *M5S3b* to contextualize real-word experience while presenting background, but unfortunately are not linguistically or rhetorically equipped to do so. In addition, interviewee D attributed the lack of *M5S3b* to the specialized content of the course.

M5S4 Expounding rationale is used to elaborate on or rephrase the theories, mechanisms, motivations, as well as common practices/knowledge related to the topic. It is often realized by conditional clauses introduced by 'if' or 'when' (see excerpt 29-32).

29) *So based on the strata, geolo- the geological time scale, GTS is a system of chronological dating that relate geological strata to time, so we can compare strata and time together, and which used by the geologist and other earth scientist. And **if** we want to describe the timing and the relationship of the event that occurred during the past. The table of geological spans agree with the nomenclature, dates and standard color codes that forthed by the International Commission on Stratigraphy. Er it was the famous commission just like the AAPG, American Association Petroleum Geologist. Ok, there are two different ways to relate time in geology, first, relative. Er, relative date just placing events in a sequence based on their position in the geological record. Er, compare with other chronologic date, Chronologic date, placing a specific number of years on an sample or rock sample. So we can see some example. Relative date, we just know the order of events but not the exactly date.* (CCL01)

In Except 29), the lecturer examined the mechanism for comparing strata and time together, as well as how the comparison could be done. The comparison was based on the definition of GTS, and specifically it was introduced by conditional clause "if we want to describe the timing and the relationship…"

30) *We can see from this picture, so the melted magma can be crystalized to be the igneous rocks. So when it can convert to the igneous rocks? So first **when** the temperature is low, cool the liquid. And second reason or second condition is increase the pressure so **when** the pressure is higher then the igneous rocks can*

convert to the. Sorry the magma can convert into the igneous rocks. And another reason, another condition is when the water, **when** *the water is removed, and that can actually raise the point, the melting point. And in that case, the magma can convert into igneous rocks. And in another way* **when** *the temperature is higher, when the pressure is lower or when the water is added into the rocks, then the igneous rocks can convert into the magma.* (CCL02)

In excerpt 30), the lecturer expounded the mechanism of inter-conversion between melted magma and igneous rocks through such conditionals as "when the temperature is low", "when the water, when the water is removed" and "when the temperature is higher, when the pressure is lower or when the water is added into the rocks".

31) *in radiation chemistry it's completely different [1.0] as the [1.3] gamma ray traverses a medium or the alpha particle whatever it is [1.0] it [0.3] interacts with the solvent [0.4] and it ionizes the solvent [1.6] you might say what happens to the solute [0.3] does it ionize that as well [0.7] well [0.3] it's purely statistical [0.7] it will [0.5] excite electrons in whatever it's passing by [1.9] and of course statistically [1.4]* **if** *you take something like methanol or pentane [0.6] then take a litre of that and work out how many m-, how many moles there are in liquid pentane [1.0] it's about ten molar [0.4] liquid pentane [0.2] is about ten moles per litre [0.8] the naphthalene is ten-to-the-minus-three [0.4] so it's a a ten-thousandfold excess of the solvent [0.7] and so ninety-nine- [0.7] point-nine-nine per cent [0.3] of all the energy [1.1] goes into the solvent [0.3]*

*and nought-point-nought-one per cent goes into solute [0.4] so [0.2] all of the
[0.5] reactions shown by the solute [1.7] reflect what's happened to the solvent
[1.6] and with this solute [0.2] it's picking up the electrons that have been
formed in the solvent [0.6] they've moved through this medium at low
temperature [0.5] and they've al-, alighted [0.3] and been trapped [0.7]
[cough] [0.2] on the solute to give this green colour [1.3] and [0.8] this process
of having a very small amount of something there [0.8] that captures [1.0] the
negative charge in the system [0.9]* (CEL01)

In excerpt 31), the lecturer explained why the interaction between gamma ray and
the solvent as well as solute is different in radiation chemistry. The lecturer further
supported the explanation by using the conditional clause "if you take something like
methanol or pentane…"

32) *okay [2.9] and [1.7]* **if** *i do the same thing for the other two [0.6] the load in
the fibre [0.3] the proportion of the load carried by the fibres [0.9] will be [0.5]
the stress in the fibre [0.5] times [0.2] the cross-sectional area of fibre [1.0]
okay [0.4] force equals [0.3] stress times area [1.1] so here P-F [0.3] becomes
sigma-F [1.0] times [0.2] P-F [1.6] and P-M here [3.9] becomes [0.2] sigma-
M [0.9] times [0.9] A-M [0.3] where [0.2] A-F [1.6] is the [0.6] total area [0.4]
or cross-sectional area of fibre [0.4] A-M is the [0.2] cross-sectional area of
matrix [0.2]* (CEL03)

In excerpt 32), the lecturer illustrated the calculation of the load in the fibre and it

is introduced by the conditional "if i do the same thing…"

Though *M5S4* is conventional in both CCL and CEL, they appear significantly more frequent in CCL than in CEL (G2= +52.71, p<0.0001). This might be due to Chinese lecturers' general instructional styles, which advocate citing authorities. All four interviewees acknowledged use of such instructional preferences. In particular, interviewees C and D mentioned that they mainly rely on authoritative resources for rationale elaboration. They seem to be comfortable delivering lectures relying on theoretic rationale proposed in or stipulated by textbooks and/or academic monographs.

M5S5 Demonstrating the topic involves explaining in detail the factual information, methods, processes, procedures, and results related to the topic under discussion, often with examples and data (see excerpt 33-38)).

33) *this kind of volcanoes are relatively lower, and often dome-like, dome-like, accumulation of basalt lava. What means basalt lava? That means contains more magnesium and iron compounds in the lava. And because the lava contains in the shallow crust, so the lave travels a long, a long distance, and spread in a thin layers. So shield volcanoes are rounded domes with gentle slopes.* (CCL05)

In excerpt 33), the lecturer stated factual information of 'shield volcanoes', which are "lower, and often dome-like, dome-like, accumulation of basalt lava", indicating more magnesium and iron compounds inside.

34) *now donors [0.5] are going to fall into two main categories those that are alive [0.6] and those that are dead [0.3] now living donors [0.3] er would may be either unrelated to the recipient [0.3] or related a sibling or or parent or whatever [1.5] obviously living donors can donate things like blood [0.4] bone marrow [0.4] and one of your two kidneys but they're going to be hard-pressed [0.3] indeed [0.2] to donate [0.3] lungs or something like that [0.4]* (CEL07)

In excerpt 34), facts on organ donors were into classified into two main categories, i.e., those that are alive and those that are dead.

35) *Let's er, let's see this the example, the world famous Panzhihua V-Ti magnetite deposit in Sichuan province, China. This deposit is very famous in the world. Em, China, was formed by crystal setting during magma differentiation. It occurred in layered gabbro body which consist of layer of mineral deposited in magma. Er, this figures show the, er, geologic profile. This is the magnetite ore. A magnetite ore is a, is the most important iron ore. Er, this is the geologic profile, this one two three, for each here. One is Syenite, two is a, a gabbro, three is light-color layered gabbro. And four to eight is layered gabbro V-Ti magnetite deposit. Here, so, the, this, this, er, magnetite deposit firstly crystalized and sink in the bottom of the magma, and then yeah here, so this, er, gabbro magnetite deposit is formed by crystal setting, here.* (CCL03)

Except 35) exhibits the process of the formation of Panzhihua V-Ti magnetite deposit in Sichuan province, China, from crystal setting to the sinking of magma.

36) *and [0.2] i suppose [0.4] the first thing that happens [0.5] is that the electrons [0.7] which are formed [0.6] in the ionization act [0.4] these are very small [0.4] and they're very mobile [1.4] and they're certainly very small compared with the [0.2] cations that are formed 'cause those are going to be molecular size [0.4] so the electrons tend to diffuse away very rapidly [0.5] simply because they are so small [0.6] and er [1.4] they have a a very high mobility [0.9] what about the er [1.0] the cations that are formed [0.4] well if you just consider a general i put R-H-plus but it could be anything [0.4] but suppose it was i don't know hexane or something [0.4] then [0.2] the [0.7] cation radical [0.4] which is the thing you get by taking the electron out of the molecule [1.0] er [0.5] will [1.4] react with electrons that are [0.2] are nearby the ones that haven't got away so to speak [0.4] and you get a certain amount of ion recombination [0.7] and when this electron returns to that cation [0.5] it's very likely [0.4] to form it in an excited state [0.2] so you can get additional excited states [0.2] from [0.3] from ion recombination [1.4] but also what can happen [0.9] is that the R-H-plus the cation radical [0.7] er is a very powerful proton donor [0.7] and it will give a proton to almost anything in sight [0.7] er and so [0.3] this will tend to happen [0.4] and perhaps the best known example of this is if you imagine [0.3] R-H-plus is [0.5] H-two-O-plus [0.6] the water cation [0.4] then [0.2] that will give away a proton to a nearby water molecule [0.4] to give H-*

three-O-plus [1.0] and you're left with O-H behind [2.2] so you do get these er

[1.6] proton transfers occurring [1.3] (CEL01)

Excerpt 36) elaborates the process and mechanism of 'proton transfer'. The process

starts with the formation of electrons and cations in the ionization act, then the two

matters react with each other to produce a certain amount of ion recombination, which

leads to additional excited states and proton transfers.

37) *Ok, let's give, let me give you for example. An example in the United States in*

nineteen ninety four, sand and gravel produced four point two six billion in

revenue, but gold produce four point one billion. (CCL03)

In excerpt 37), the lecturer used the revenue data of United States to show the

importance of mineral resources.

38) *here are some figures for [0.6] southern Africa [0.9] looking at prevalence*

rates [0.5] in different [0.3] states different provinces of southern Africa [1.1]

and look at the tracking since the early nineteen-nineties since nineteen-ninety

[0.7] through to nineteen-ninety-seven [1.7] just look how fast [0.7] the

epidemic is spreading [1.1] you've gone from [0.4] you know [0.2] about one

per cent one or two per cent [0.3] in Kwazulu and Natal [0.7] in nineteen-ninety

one to two per cent of adults being infected which itself is bad enough [0.7] to

now [0.5] greater than twenty-five per cent [0.4] adults infected [0.4] in that

small province [4.3]similar figures for the s-, Soviet Union [2.2] newly number

of diagnosed infections [0.3] starting in nineteen-eighty-seven [0.7] and running through to nineteen-ninety-seven [0.5] (CEL06)

Excerpt 38) reveals figures on the prevalence rates of the AIDS epidemic in southern Africa.

As the most frequent and essential step in academic lectures, **M5S5** is conventional in both CCL and CEL. The overwhelming frequency could justify the lack of significant difference (G2= +1.36, p>0.05) of this step between the two corpora.

M5S6 Providing caveats is used to explicitly articulate caveats related to the topic. Caveats are important message essential for proper understanding of the topic, but they tend to be neglected for various reasons. **M5S6** is often realized by the use of such verbs and auxiliaries as 'remember', 'note', 'need', 'have to', etc.

39) *Then **remember** also er, not all the fracture material can be universal, for each reservoir they own, they have own like er application range. You cannot say ok I I design one fracture fluid I can use anywhere. There is no such thing. So you need put also keep this thing in your mind. You need said ok er this one work well. This one ok I cannot take it to another reservoir. Then you need redesign it. And then you can use it.* (CCL08)

40) *er do **note** that when you've got a system like that when you're in second gear or third gear or fourth gear [0.7] the effective [0.8] moment of inertia of that system [0.5] becomes greater [0.6] as you go up through the gears [2.4] that in fact is quite [0.2] a a complicated thing [0.3] to work out and you have to know*

something about gear boxes and gear ratios [0.2] but we're not concerned with

that here [2.2] (CEL02)

In excerpt 39), the lecturer reminded students of the application range of fracture material whereas the lecturer in excerpt 40) warned students of the complication of inertia brought by gear boxes and gear ratios.

M5S6 appeared in 25% (3/12) and 83% (10/12) of lectures in CCL and CEL respectively, thus is optional in CCL, but conventional in CEL. Log-likelihood comparison also reveals significantly less ($G2$=-8.03, $p<0.01$) use of this step by native Chinese lecturers than native English lecturers. This could be an indication of difference in terms of teaching goals and preferences between lecturers in CCL and CEL. All four interviewees contended it is essential to use ***M5S6*** to provide caveats for better comprehension and mastery of the lecture content. However, some Chinese EMI lecturers, as representatives of the non-native English lecturer group, may have neglected it, or perhaps may be linguistically incapable of doing so.

M5S7 Making comments allows the lecturer to share with students his/her understanding of and comments on the topic. This step is usually realized through the use of adjectives, such as 'important', 'key', 'critical', comparative/superlative adjective forms, or stance verbs.

41) *Well, er, critical velocity are different for sediment entrainment and deposition,*

*especially in the finer fractions. Er, the velocity is very, er, very **critical** to them.*

*And fluid density and velocity, er, viscosity play a **key** role in determining which*

one can transport, the density and viscosity, er, viscosity, are the key, are the

keys to determine which one, the big one or the small one, or the medium, or

the middle one. (CCL04)

42) *so this is generally regarded as the **best** solution it's not always [0.3] a*

practical solution particularly if you're [0.5] er in a rocky area and need to

[laugh] [0.4] dig a dig a deep pit to do this [1.8] and it also has to be looked

after [2.5] you have to er make sure this pit is kept [0.2] kept clear and weed

free [1.1] (CEL04)

In excerpt 41), the lecturer commented on the key role of fluid density and velocity, and in excerpt 42), the lecturer regarded rainfall measurement technique as best solution, though sometimes not practical, especially considering the pit-digging and maintenance in rocky areas.

M5S7 appeared in 58% (7/12) and 100% (12/12) of CCL and CEL lectures respectively, thus is conventional in both, indicating the importance of such a step in academic lecture delivery. However, significant difference (G2= -9.47, p<0.01) was found in the use of this step between CCL and CEL lectures, with the Chinese lecturers using significantly less of it.

Interview data showed split preferences among the four interviewee lecturers. Interviewees A and D affirmed their tendency to use *M5S7* to share with students their understanding of and comments on the topic. However, Interviewee B insisted avoiding comments since they may sometimes touch upon touchy issues and cause

misunderstanding due to different cultural values. Interviewee C said she would not make comments since she would prefer to demonstrate facts and respect results.

On the other hand, native English lecturers seem to prefer sharing their comments, so as to influence students with their critical thinking.

M5S8 Summing up the topic serves to summarize the topic to consolidate learning of students. This step is often explicitly indicated by verbs of summary, such as 'summarize', 'review', etc.

43) *So let's **review** what we said, igneous rocks. What is igneous rocks? It is melted, it is formed when the melted rocks cools, and solidified. That's how the igneous rocks is formed and what is igneous rocks. And second is formed from magma beneath the crust or the lava above, above the crust, exposed outside. So the igneous rocks can be formed from two things, magma inside or lava above. And the crystal size of the igneous rocks can be formed, can be determined, can be determined where it is formed, where it is formed. And based on the crystal size and where it is formed, we can actually divide, classified the igneous rocks as like this, we can see later. So the igneous rocks can be classificated, can be classificated where they were formed, and besides this, the igneous rocks can be also calssificated by the compositions of them, so which mineral inside and how many minerals inside.* (CCL02)

44) *okay [0.4] so that **summarizes** these points really [1.7] in the absence of immunosuppression grafts are [0.3] inevitably rejected [1.3] acute rejection*

occurs [0.9] after a delay i suppose because in the presence of immunosuppression it takes a long time [0.4] for the immune responses to develop [0.4] something that w-, should normally take a few days [0.3] takes a few weeks because you're reducing the proliferation of the cells [2.7] and that's what i'm saying in this last point that anti-graft T-cells grow more slowly under immuno-, immunosuppression [3.8] (CEL07)

In excerpt 43), the lecturer summarized what she had taught about the formation and classification of igneous rocks through "*let's review…*", while the lecturer in excerpt 44) summarized the reaction of T-cells to grafts in the presence and absence of immunosuppression.

M5S8 appeared in 42% (5/12) and 75% (9/12) of CCL and CEL lectures respectively, thus is optional in CCL, but conventional in CEL. No significant difference (G2= +0.17, p>0.05) was found between the two corpora in the use of this step.

M5S9 *Initiating co-building lecture* functions to engage students in co-building the lecture, thus facilitating learning among students. This step involves questions raised and answered by both teachers and students to co-build the lecture so as to enhance students' learning. It generally consists of several rounds of turn-taking between teachers and students (see excerpt 45-46)).

45) *S1: ...Ok. So I have a question. Er for bubble flow and annular flow, which one has- which one's gas have a bigger velocity? Which one? The gas has a bigger velocity? Which one?*

SS: The annular flow.

S1: Yeah, the annular flow, right, annular flow. Ok. Because for this you can see, here we have a small blue, blue bubbles, right? And others is water, or oil, that's liquid, ok. And for this flow, you can see, the liquid is flow along the wellbore, right? Along the tubing. And the gas flow in the middle of tubing, ok. And from this flow regime to this flow regime, the gas velocity is increasing, ok, er, is increasing, er. So this is four flow regime along the tubing, er, along the tubing.

(CCL07)

In excerpt 45), the lecturer explicitly asked whether bubble flow or annular flow has a bigger velocity to check students' understanding of the topic. Then students offered their answers. And the lecturer followed by giving detailed explanation of the answer, thus students' learning was, more or less, consolidated.

46) *nm0881: ... the [0.4] the main type of rain gauge are storage gauges [1.6] where [0.3] these are generally read [1.0] these are s-, [0.2] simple [2.5] er collectors [0.3] of rainfall [1.2] and they're normally just read [0.9] er once per day [7.3] matters up to even at operational [0.5] weather sites [1.0]*

sm0882: sorry i

nm0881: that that you that you measure only once per day from a rain gauge [0.4] i mean certainly on a site like ours [0.2] they're only read once a day [0.5]

sm0882: yeah you you [0.2] you then for the standard five inch gauges but the

nm0881: mm

sm0882: there are [0.9] automatic loggers which actually do what the

nm0881: yeah well i understand you [0.4] so these these are what so so at some sites they would be measured twice a day but say on our [0.3] our climatological web [0.6] climatological site [0.4] we'll only use them once a day [0.5]

(CEL04)

Excerpt 46) consists of several rounds of turn-taking. Different from excerpt 45) where the lecturer initiated the comprehension-checking question, the student cut in by saying "*sorry I*" and "*yeah you you [0.2] you then for the standard five inch gauges but the*' and '*there are [0.9] automatic loggers which actually do what the*". The lecturer responded to the student's doubt about the times of rain gauge measurements. In this way, the co-building of the lecture was also realized, only that it was started by the student.

M5S9 appeared in 67% (8/12) and 100% (9/12) of CCL and CEL lectures respectively, thus is conventional in both CCL and CEL. No significant difference (G2= +0.44, p>0.05) was found in the use of this step between the two corpora. This seems suggest that both native Chinese and native English lecturers pay heed to involving students so as to improve lecturing effectiveness.

M5S10 Pinning down and/or clarifying the topic is used to confirm and clarify the topic. This step is often realized through the use of such expressions as 'What I want to say', and 'what I'm doing', etc., for clarification of the topic being discussed (see excerpt 47-48)).

47) *And here, a one, b one is des-, is described as this formulas, oh this formulas.*

What I want to say*, because the parameters here is the common parameters we have, have been seen, so here I will not, I will not want to er explain again. And what I want to say is that here the a one is the intercept of the formulas, of the type A waterflooding curve, and b one is the slope of low line.* (CCL09)

In excerpt 47), the lecturer articulates that such parameters as '*a one*' and '*b one*' are common in the area of waterflooding, thus would not be explained further.

48) *so [0.2]* ***what i'm doing*** *here is taking [0.8] a moment [1.8] of the impulses P and R about the centre of gravity i'm working at the centre of gravity [0.6] so P [1.8] gives you [0.4] er a cl-, [0.3] a clockwise moment of impulse of P times the distance H-minus-L [0.8] and the R gives you [0.5] a [0.2] also a clockwise [0.4] moment of impulse [0.5] of R times L [1.1] so you've got these two terms [4.2] in the expression for the moment of impulse [1.0] and that [1.0] will result in a change in angular momentum [1.0] I-theta-dot [0.6]* (CEL02)

In excerpt 48), the lecturer clarified the two terms '*moment of impulse of P*' and '*moment of impulse of R*' that were used in the expression for the moment of impulse.

M5S10 appeared in 17% (2/12) and 92% (11/12) of CCL and CEL lectures

respectively, thus is optional in CCL, but conventional in CEL. There are significantly less (G2= -19.54, p<0.00001) *M5S10* identified in CCL than in CEL.

The interview data deviate a bit from text analysis results of the lecture manuscripts. All four interviewees insisted the necessity of using *M5S10* to clarify topics in accordance with students' responses and difficulty level of lecture content. This might be related to the emerging status of EMI courses in China. Chinese lecturers fully understand the importance of topic clarification, but might be less experienced in doing so in a foreign language other than their native language. On the other hand, native English lecturers seem to be more comfortable with and experienced in using *M5S10* to clarify their instructions so as to ensure and improve the lecture effectiveness.

M6 Building Theme Network functions to help build a whole knowledge network that may serve the lecture theme. This move connects topic nodes in the theme network so as to help build the knowledge system of the course. This move is realized through three different steps, viz., *M6S1 Referring to previous lecture/other source*, *M6S2 Specifying subsequent/future content* and *M6S3 Connecting/Comparing topics*. Generally, there are significantly less (G2= -7.76, p<0.01) use of *M6* by lecturers in CCL than in CEL. *M6* appeared in 92% (11/12) and 100% (12/12) of lectures in CCL and CEL respectively, thus are conventional in both corpora. A closer scrutiny of the constitutional steps revealed that the significant difference was mainly due to the different use of *M6S1*.

M6S1 Referring to previous lecture (part)/other source involves creating

knowledge network by connecting the present lecture with previous lecture (parts) or knowledge from other sources. It makes connections between earlier part and present part of the present lecture, between the present lecture and the previous ones, as well as contents from other sources such as textbooks, research articles or academic monographs in the disciplinary area (see excerpt 49-52)).

49) *We already, we've already studied about it before. And when these kinds of rocks in mantle or upper mantle or lower crust, when it's melted, then the magma is formed. And we also mentioned another content, that is concept, that is lava. So lava is a kind of special magma when it's exposed outside of the crust.* (CCL02)

In excerpt 49), the lecturer reminded students that they have dealt with the concept of magma, rocks in upper mantle or lower crust. In addition, they have also mentioned lave, a kind of special magma outside of the crust. Putting together the contents covered before, a theme network of igneous rocks was created.

50) *and let's actually go through this algorithm again, we said there were N minus one passes, here N is five so we'll do we'll do four passes, and we said each one, uh essentially takes us through the array elements one time, we look at them all, we take the, item that's in the starting position for the current pass, and we exchange that or swap it, with the smallest item, that's in the unsorted part for the current pass. so essentially the idea is on one pass to sort one item, and we know from yesterday's, fairly quick discussion that after four passes,*

five elements would be sorted. and that's always the case, always N minus one.

(CEL12)

In excerpt 50), the lecturer went through the algorithm of yesterday's discussion that it was always the case that after N minus one passes, N elements would be sorted. The lecturer set the scene to connect the algorithm discussed yesterday to a new algorithm to be covered today.

51) *And we use the technology we call it Nodal analysis, ok, or Nodal analysis. This is Schlumburger's patent, ok. You know Shclumburger is a very famous international oil company, right? And here why we use Nodal analysis? Er, why, we have three answers, ok. The first one the fluid properties change with the location-dependent pressure and temperature in the oil or gas production system, right? Er, when the pressure changes, when the temperature changes, the fluid properties will also change, right? And second, the node can break the system into discrete, er, into, into several element, ok, er, several element. So you set a node in the system, you can separate the proce- the process into several parts. And several parts have the, its even properties, ok. And the third, fluid properties in the element are evaluated locally. Ok, so this is three different reasons for why we use Nodal analysis.* (CCL07)

In excerpt 51), the lecturer explained in detail why the '*Nodal analysis'* technology, knowledge of other source, was adopted to predict the well deliverability. Therefore, a theme network was built between the '*Nodal analysis' technology of Shclumburger* and

the present lecture.

> 52) *some of the material not all of it, comes from a paper, that uh White and John White and i published in nineteen eighty-four so it's been a while, uh this is not all of the material mind you okay so there's some extensions that you're gonna see in class, which are not shown in the paper, but some of the basic derivations are here, this is the... <WRITING ON BOARD> Bozer and White paper...I-I-E transactions, i won't write the title but it's called Travel Time Models for Automated Storage Retrieval Systems... volume... well i may not have the volume number here... yeah i just_ yeah volume sixteen number four... and as i said it's a fairly old paper nineteen, eighty-four,* (CEL10)

In excerpt 52), the lecturer made it clear that some of the class material was extensions to the paper co-authored by John White and the lecturer, and the lecturer also referred to the Bozer and White paper while explaining the derivation of the cycle time for unit load automated storage retrieval systems.

M6S1 appeared in 67% (8/12) and 100% (12/12) of lectures in CCL and CEL respectively, thus is conventional in both corpora. However, significant differences ($G2= -16.60$, $p<0.00001$) were found between lectures in the two corpora. Chinese lecturers used significantly less *M6S1* than their English counterparts. This could be, on the one hand, related to the course syllabus and delivery of the CCL lectures as mentioned in the discussion of *M2S2*. The twelve CCL lectures selected from the four courses were all co-taught by different lecturers, therefore may, to some extent,

undermine the basis for *M6S1*. On the other hand, the less use of *M6S1* could be attributed to Chinese EMI lecturers' less experience in building knowledge connections through this step though interview data revealed unanimous agreement on the importance of it.

Meanwhile, though CCL lectures were co-taught by different lecturers, they generally last for 16 weeks, which brought about certain levels of continuity in terms of both lecturers and lecture contents, hence conventional status of *M6S1* in CCL. This is in alignment with Lee (2009), which identified a similar step as obligatory. Nonetheless, Thompson (1994) found that this step appeared only in 33% (6/18) of the lectures. This might be related to the lectures under investigation. Among the 18 lectures of Thompson (1994), nine were 'one-offs', which may, to a great extent, justify the lower frequency of this step in her study.

M6S2 Specifying subsequent/future content is used to reserve a node in the knowledge network by specifying subsequent/future content that may support the theme in the present lecture. This step is usually indicated by verb phrases and time phrases showing plans of the future, i.e., 'we are going to', 'we'll', 'in section three' and 'later' (see excerpt 53-54)).

53) *So this is technology probably **we are going to** talk **in section three***. *I think **we're going to** talk about that*. (CCL08)

54) *again which **we'll** [0.3] come to **later** [0.4] **in the lecture course** [0.3] which effectively just measures the proportion of the variation or tries to [1.0] of the*

total variation that's genetic [0.7] so it gives us some sort of estimate of how

efficient [0.3] or how effective our selection's going to be [2.2] [cough] [4.1]

it's something i'll talk about when we talk about heritability [0.3] (CEL05)

In excerpt 53), the lecturer specified that he would talk about the technology later in section three, so students may try to reserve a node connecting the present content with the future content. In excerpt 54), the lecturer announced that they would deal with genetic variation later in the lecture course when they talk about heritability, thus a connection between the present lecture content and that of the future could be reserved.

M6S2 appeared in 33% (4/12) and 67% (8/12) of lectures in CCL and CEL respectively, thus is optional in CCL, but conventional in CEL. No significant difference (G2= -3.53, p>0.05) was found between lectures in the two corpora. Again, given the teaching syllabus and delivery of lectures in CCL, less continuity of lectures in CCL is expected, thereby less *M6S2*.

M6S3 Connecting/Comparing topics is used to connect or compare topics. This step is often realized through comparative adjective phrases and/or adjectives of comparison/contrast. *M6S3* is different from the deceptively-semblable step *Relate 'new' to 'given'* identified by Thompson (1994) and Lee (2009) because the purpose of *M6S3* is to make connections and comparisons of topics for deeper understanding among the students whereas the purpose of *Relate 'new' to 'given'* is to put the present lecture in the given context of shared knowledge or experience (see excerpt 55-56)).

55) *But the **different** to gabbro is, we can see, the color is **much lighter**, right?*
The color is much light than the gabbro. That's actually because they have a
different** kind of composition to the gabbros they have the **less** magnesium, **less
*iron, but **more** silica, more silica. So we can see even they have, they, I mean*
the intrusive igneous rocks, they are from the same things magma, but because
*they have the **different** compositions, then they, they have **different** kind of*
characteristics (CCL02)

In excerpt 55), the lecturer compared the characteristics of diorite with gabbro, and
found that the color of diorite is much lighter due to less magnesium, less iron and more
silica components contained. In this way, a connection was built between the two topics,
so that students may have a better chance of building the theme network of the lecture.

56) *the the key thing here [0.3] is to see the **parallel** [0.4] here between linear [0.3]*
*particle motion [0.3] and rotational motion [0.7] and we did a very **similar***
thing for that trivial problem to do with [0.3] railway trucks colliding with each
other if you remember it [2.3] all right moving [0.5] on then [1.2] just in passing
[0.3] there is a chart at the back of your notes showing er [0.4] the comparison
between [0.6] translational motion [0.3] and [1.1] rotational motion [0.4]
(CEL02)

In excerpt 56), in order to explain the differences between linear motion and
rotational motion, the lecturer referred to the example of "railway trucks colliding with
each other". Through the analogy of the railway trucks, the lecturer made clear the

differences. Such analogous comparison and connection expect to help students to understand the topic better.

M6S3 appeared in 92% (11/12) in both CCL and CEL lectures, therefore is conventional in both corpora. Nevertheless, no significant difference (G2= +0.50, p>0.05) was found between the lectures in the two corpora. This might be an indication of the general importance of *M6S3* in academic discipline lectures.

In brief, despite different use of *M6S1*, both native Chinese and native English EMI lecturers recognize the importance of building theme network through connections between knowledge points.

M7 Making Aside refers to classroom talk less germane to lecture content, but may be conducive to understanding the topic, as well as building and maintaining good teacher-student rapport. Essentially, aside is peripherally related to the lecture contents, but it does not include any small talk that is used to buffer the serious lecture content like that of the digression step identified by Lee (2009, 2011, 2016) in lecture introductions (see excerpt 57-58)).

57) *S1: ...Jurassic and Cretaceous er, we, I guess most of us see the movie.*

S6: Jurassic World.

S1: Very famous and serious movie.

(CCL01)

The communicative purpose of excerpt 57) is not only to build rapport with students, but the talk of the movie '*Jurassic World'* is intended to arouse students'

interests and imagination of the geologic time, which was the lecture theme.

> 58) *uh now i drew this by hand because i was having trouble getting my, uh, graphing calculator to allow me to take, the image it produced for the graph and then put it in the power point show. but hopefully this is a good enough, approximation that you can see, how the processing time increases.* (CEL12)

The purpose of 58) is not just to inform students of the lecturer's trouble with the graphing calculator. Instead, the lecturer intended to get students to understand the increment of processing time through hand-drawn graphs. To a great extent, the asides lecturers made were, in one way or another, related to the topics in the lecture.

M7 appeared in 17% (2/12) and 75% (9/12) of lectures in CCL and CEL respectively, therefore is optional in CCL, but conventional in CEL. However, no significant difference (G2= -3.51, p>0.05) was found between lectures in the two corpora. One possible reason could be the relative low frequency of *M7* in the two corpora (4 occurrences in CCL and 19 occurrences in CEL).

M8 Housekeeping is used to give recommendations and offer reminders. Functionally, it is similar to *M2S4 Housekeeping*. However, *M8* generally appeared in the middle or at the end of a lecture, while *M2S4 Housekeeping* is generally used at the Opening Phase of a lecture to make class management announcements and offer reminders.

59) *can i remind you that we've got [0.4] the tutorial tomorrow [0.6] okay [0.3] whatever is the tutorial time slot i think it's eleven to twelve [0.5] in G-twenty-six [0.4] okay [1.7]* (CEL03)

60) *Also there are some, there is some useful links. Er, if you are interested, you can go to this website to find more detailed information about the exogenti-exogenetic or endogenetic processes.* (CCL05)

61) *and i'm not going to [0.3] to do any more and talk about this if you're interested in them then both [0.3] er [0.7] two of the books that i i've referred to [0.2] Ward and Robinson [2.0] and Strangeways [2.0] er go into quite a lot of detail about snow measuring techniques it's really quite an interesting area [0.7] and of course if in some [0.5] continelt-, ar-, continental areas [0.4] it's an important contribution to the er [2.2] to the whole hydrological cycle [6.3]* (CEL04)

In excerpt 59), the lecturer reminded students of the next-day tutorial, the function being similar to that in *M2S4*. However, in excerpt 60), the lecturer recommended that interested students should go to the website to find detailed information on endogenetic processes. And in excerpt 61), the lecturer suggested students read the two books by Ward and Robinson and Strangeways, respectively.

M8 appeared in 25% (3/12) and 83% (10/12) of lectures in CCL and CEL respectively, therefore is optional in CCL, but conventional in CEL. There were significantly less (G2= -19.00, p<0.00001) *M8* used in CCL lectures than in CEL

lectures. Interview data revealed different attitudes among Chinese EMI lecturers. Interviewee C affirmed her own preference of reading material recommendation through *M8*; interviewee A argued for reading recommendation in accordance with lecture content; interviewees B denied the necessity of reading recommendation given the comprehensiveness of the textbook; interviewee D deemed it unnecessary to do the recommendation on account of poor professional background and performance of the students in his lecture.

In general, different attitudes towards textbooks and different classroom practices between Chinese and English lecturers may have contributed to different use of *M8*. Though it is not common practice for Chinese EMI lecturers to give recommendations or offer reminders through *M8* while lecturing, native English lecturers seem to prefer doing these. These strategies could, to some extent, help improve their teaching.

M9 Checking Comprehension & Consolidating Learning serves to check on students' comprehension as well as to consolidate learning. This step is generally realized by a combination of procedural questions on classroom teaching and management, rhetorical questions on the lecture content raised and answered by teachers, and genuine questions on the lecture content raised by teachers/students but answered by individual/all student(s) (See excerpt 62-66)). All are for purpose of checking understanding and consolidating learning. Since checking understanding and consolidating learning are essentially intertwining, it does not make sense to impose an artificial dichotomy to split it into two individual move/step. The present study

combined the two concepts.

62) *Do you know understand it?* (CCL01)

63) *okay...? any other questions? <PAUSE:08>* (CEL10)

In excerpts 62) and 63), the lecturers raised the procedural questions to check students' understanding. They usually function to ensure the instruction is clear, so the lecturer could proceed with lecturing. Generally, there is an elapse of silence or a pause, indicating no question from the students.

64) *Ok, let's see some questions. Which metal is frequently mined in placer deposit? A copper, iron, lead or gold? Gold, here, yeah. We just learn that the gold will be formed this place, placer deposit yeah. Which metal ore deposit can be formed by chemical precipitation? Er, yeah, maybe here, maybe next class we will learn this, yeah. Let me some figure, yeah, ok. Er, let this, this question. Most of the metal ore deposit was formed by?*

SS: <murmur>

S1: Magmatic-hydrothermal fluid, yeah. Yeah, which one is not energy resources? Yeah, this is the metal. Yeah, and most of the metallic ore deposits are associated with? With what? With what? Igneous rock, yeah. Ok. Ok, this question and homework, you can write down. OK. (CCL03)

In excerpt 64), the lecturer raised a number of questions concerning the components of ore deposit, which he himself provided answers subsequently as a strategy to enhance learning of the students.

65) *So this is why we inject gas can help us to lift the oil, why?*

S3: Because we still have oil, so the only way we have to inject the gas to flow to effective density, so it can flow.

S1: Yeah, that's right. That's right. So the gas we inject to the annulus, and then it will flow into the tubing, right? When the gas flow into the tubing, the fluid density will be decrease, er, fluid density will be decrease. So the decrease density is easier to be lift, right? Er, this is why we use gas lift. (CCL07)

In excerpt 65), the lecturer raised the question about the reasons to inject gas to help lift oil, which was replied by one student. Then the lecturer affirmed the student's answer and proceeded to add more explanation. It is through this Initiation-Response-Feedback pattern of questions that the lecturer and students achieve learning consolidation for the whole class.

66) *S1: ...Mary did you have a question?*

S6: so you're saying that, the need to use a regulatory system, increases your standard metabolism? S1: um, typically a change in the environment from the preferred environment, can be thought of as an addition to standard metabolism. it increases the energy required just to keep body and soul together. a_ for you, in winter, with inadequate s- s- central heating, you will, make up for the inadequate central heating, as so often happens in campus housing i understand, uh will be maki- make up for that, by increasing your metabolism, if it really gets cold you'll start shivering, but you'll still crank up your

metabolism a little. on the other hand in um uh, in in the summer, when you'll

be sweating like a pig all the energy will be required for that, um, if you don't

have air conditioning. and these can be seen as an increase in the minimum

energy the standard metabolic rate, in order to keep body and soul together.

(CEL08)

In excerpt 66), the lecturer noticed that Mary may have problems understanding the lecture, so the lecture asked about Mary's question. Then Mary asked the lecturer to clarify why the regulatory system is needed to increase standard metabolism. Subsequently, the lecturer offered detailed explanation on the reasons. This could be an effective way to promote understanding of the lecture content for students.

M9 appeared in 42% (5/12) and 50% (6/12) of lectures in CCL and CEL respectively, therefore is optional in CCL, but conventional in CEL. There is significantly less (G2= -3.93, p<0.05) *M9* found in CCL than in CEL. Interview data revealed that all interviewees acknowledged the importance of checking students' comprehension and consolidating learning. Nevertheless, they all prefer to do so through assigning homework, instead of verbal checking on site. It is perhaps not common practice for Chinese lecturers to verbally check students' understanding in class, as indicated by interviewee C, who articulated that "even if I asked whether they have any questions, it was meant to be procedural or ritual". On the contrary, native English lecturers seem to be more comfortable using this strategy to help improve classroom teaching effectiveness.

4.4 Rhetorical moves in the Closing Phase of EMI lectures in CCL and CEL

The Closing Phase serves to summarize the lecture, get students to cool down and say farewell. It consists of three moves with different steps.

Table 4.4 Moves/steps in the Closing Phase

Moves	Distribution [a]		Frequency [b]		LL critical value G2 [c]
	CCL	CEL	CCL	CEL	
M10 Wrapping Up Lecture	7 (58%) [#]	2 (17%)	8	2	+ 8.51**
M11 Cooling Down	8 (67%) [#]	10 (83%) [#]	13	19	+ 0.44
M11S1 Housekeeping	5 (42%)	5 (42%)	6	7	+ 0.68
M11S2 Looking ahead	5 (42%)	9 (75%) [#]	7	12	+ 0.03
M12 Ending Lecture	12 (100%) [#]	12 (100%) [#]	15	16	+ 2.34
M12S1 Dismissing class	8 (67%) [#]	10 (83%) [#]	8	12	+ 0.22
M12S2 Farewell	7(58%) [#]	3 (25%)	7	4	+ 3.72

Note. [a] Distribution refers to the number of lectures where certain moves/steps are present. [b] Frequency refers to the number of moves/steps that are identified. [c] The log-likelihood (LL) critical value G2 itself is always a positive number. However, indicators '+' and '-' are inserted to show overuse and underuse of corpus 1 (in this case CCL) relative to corpus 2 (in this case CEL). *$p < 0.05$, critical value G2= 3.84; **$p < 0.01$, critical value G2 = 6.63; ***$p < 0.001$, critical value G2 = 10.83; ****$p < 0.0001$, critical value G2 = 15.13. [#] indicates conventional status of moves/steps.

Table 4.4 provides distribution and frequency data of each move/step in the Closing Phase.

All three moves appear in both CCL and CEL. According to the conventionality criterion stipulated in the present study, all the three moves are conventional in CCL, *M11* and *M12* are conventional in CEL, but *M10* is optional in CEL. This seems to indicate similar rhetorical move structure in the Closing Phase of the two corpora.

The three moves in the Closing Phase are different from those moves identified by

Lee (2011, 2016), viz., *M7: Setting Up Homework Framework*, *M8: Cooling Down*, and *M9: Farewell*. The major difference lies in *M10 Wrapping Up Lecture* identified in the present study. The language-orientation of the EAP lectures in Lee (2011, 2016) might be the reason why no such move was identified in his study. The present study involves academic lectures, which are topic-oriented, thus lecturers might be inclined to include *M10* in lecturing.

The homework related move *M7: Setting Up Homework Framework* identified by Lee (2011, 2016) was partially embodied in *M11S1 Housekeeping* of the present study, though the steps of *homework outlining*, *modeling* and *checking* of Lee (2011, 2016) was nowhere to be found in the present study. Again, this could be also due to the different lecture orientations between Lee (2011, 2016) and the present study.

M10 Wrapping Up Lecture is used to make summaries of the lecture content. It is realized by listing the major contents of the lecture. One conspicuous difference between the CCL and CEL lectures lies in comments on the summary of lecture contents. Chinese lecturers tend to only list the major contents as in excerpt 67), while English lecturers prefer to provide further comments on these lecture contents 68).

67) *Ok, now, this class, we have mainly talked about what is waterflooding curve, Ok? You fir-, you should first know this. Ok two related parameters, and we plotted it on the coordinate, we can find the, we can find the straight line. Ok this is the waterflooding curve. The second is the characteristics of three typical waterflooding curve and their relationships. We have told three typical*

waterflooding curve A, B, C. They are parallel to, A and B, they are parallel to each other, but their intercept is different, right? And the third, we have talked about the application of the waterflooding curve. Yes, the first we can determine the recovery, the recoverable reserve and recovery. The second we can predict the dynamics. The third we can, we can calculate the dynamic reserve, right? Ok, this is for our class <POINTING AT THE PROJECTOR SCREEN>. Er this is for our class. (CCL09)

68) *so we've talked about [1.2] radiolysis [0.3] using scavengers of organic systems [0.3] and i've finished up by talking a bit about water [0.5] the big difference is the organic systems [0.2] have excited states that's a very important part of their radiolysis [0.8] water [0.5] is entirely [0.4] ionization [0.3] to give [0.2] O-H-dot [0.4] and E-minus those are the two principal species [0.3] and they dominate radiation biology [0.9]* (CEL01)

M10 appeared in 58% (7/12) and 17% (2/12) of lectures in CCL and CEL respectively, therefore is conventional in CCL, but optional in CEL. There is significantly more (G2= +8.51, p<0.01) *M10* found in CCL than in CEL. All four interviewees contended that they always made lecture summaries before ending the lecture, though it might be dependent on the lecture time left, as indicated by interviewee A. This practice could be related to the fact that Chinese lecturers tend to provide students with definite conclusions due to their preferences to inductive thinking, while native English lecturers may opt to leave open for students to explore more by

themselves.

M11 Cooling Down serves to attend to course-related matters and/or to discuss future lectures. It consists of two steps, viz., ***M11S1 Housekeeping*** and ***M11S2 Looking ahead***. ***M11*** appeared in 67% (8/12) and 83% (10/12) of lectures in CCL and CEL, thus is conventional in both corpora. No significant difference (G2= +0.44, p>0.05) was found between lectures in CCL and CEL. This might be due to consensual attitudes towards the importance of ***M11 Cooling Down*** by both Chinese and English lecturers.

M11S1 Housekeeping is used to assign homework, and offer homework-related reminders.

69) *And the last part about the partial melting, and I think I won't talk about this here and that's kind of your homework. You give a research, research some information about the partial melt and research some information to know what's the relationship between the partial melting and the different kind of igneous rocks is formed.* (CCL02)

70) *there's a a lot of material in here, and, it does require a mindset which is a little different, from the one you're used to. e- the mindset you're used to in science classes is, measurable quantities using human technology. this is a mindset, that tries to take an animal eye response-related view, of the world. so it's a little different, and my experience has been, that you probably need to think a little about this. and of course we'll have a review session later.* (CEL08)

In excerpt 69), the lecturer assigned homework verbatim to students by saying '*that's kind of your homework*', whereas lecturer in excerpt 70) reminded students of required mindset for processing large quantities of material while doing homework.

M11S1 appeared in 42% (5/12) of lectures in both CCL and CEL, therefore is optional in both corpora. However, this is different from Lee (2011, 2016), where the *M7S1: Announcing homework* is conventional. This might be due to different viewpoints towards homework by discipline lecturers and language lecturers. No significant difference (G2= +0.68, p>0.05) was found in the use of *M11S1* between CCL and CEL.

M11S2 Looking ahead presents students a preview of upcoming lectures. Through *M11S2*, lecturers inform students of the forthcoming lecture themes in advance.

71) *And next class we will talk about another empirical method, that is the decline, decline analysis,* (CCL09)

72) *next week we'll talk about uh human populations and, allometry.* *<UNINTELLIGIBLE SPEECH>* (CEL11)

M11S2 appeared in 42% (5/12) and 75% (9/12) of lectures in CCL and CEL respectively, therefore is optional in CCL but conventional CEL. Lee (2011, 2016) also identified the optional status of the similar step *M8S1: Looking ahead*. This seems to indicate that EMI lectures in CCL are prone to the language-oriented EAP lectures of native English lecturers. No significant difference (G2= +0.03, p>0.05) was found in the use of *M11S2* between CCL and CEL.

As the last move of a lecture, ***M12 Ending Lecture*** declares the ending of the lecture. It consists of two steps, viz., ***M12S1 Dismissing class*** and ***M12S2 Farewell***. It is a conventional move in both CCL and CEL, since it occurred in all lectures of the two corpora. No significant difference (G2= +2.34, p>0.05) was found, indicating consensual opinion of the necessity of this move by Chinese and English lecturers.

M12S1 Dismissing class signals the ending of the lecture.

73) *Ok. That's all today.* (CCL12)

74) kay well it's after ten so we'll stop here, (CEL12)

In excerpts 73) and 74), the lecturers explicitly announced the ending of the lecture. ***M12S1*** appeared in 67% (8/12) and 83% (10/12) of lectures in CCL and CEL, thus are conventional in both corpora. Log-likelihood comparison revealed no significant difference (G2= +0.22, p>0.05) in the use of ***M12S1*** between lectures in CCL and CEL.

M12S2 Farewell ends the lecture with farewell and/or gratitude.

75) *see you nine o'clock tomorrow morning* (CEL04)

76) *Ok, thank you for your attention. Ok.* (CCL08)

77) *thanks very much indeed* (CEL06)

Excerpt 75) is an example of ending the lecture through ***M12S2*** by the lecturer. But more often ***M12S2*** was realized by showing gratitude to students' attention and cooperation, as in excerpt 76) and 77).

M12S2 appeared in 58% (7/12) and 25% (3/12) of lectures in CCL and CEL, thus is conventional in CCL, but optional in CEL. However, no significant difference (G2=

+3.72, p>0.05) was found in the use of *M12S2* in lectures between CCL and CEL.

4.5 An ELF perspective of rhetorical moves in EMI lectures

Lorés-Sanz (2016) examined move simplification and hybridity in the abstract writing of native and non-native academics in Sociology from an ELF perspective. Analogically, the present study examines whether such move simplification and/or hybridity exist among Chinese EMI lecturers in contrast to their native English counterparts.

As suggested in Section 4.1, no significant difference was found in terms of the frequency of moves/steps in EMI lectures given by native Chinese and native English lecturers, which denies any move simplification in CCL lectures. Interview data have revealed the existence of genre transfer in EMI lectures from their Chinese lectures, thus there is no need for such move simplification. Therefore, the present study only investigates the move hybridity, if any, in EMI lectures given by Chinese lecturers. This is different from the evolving process in academic writing, which could take dozens of years for hybridity of move structure to occur as suggested by Lorés-Sanz (2016).

However, no attempt was made to determine a sequential order of moves and steps due to the frequent online processing involved in lectures. In addition, the punctuations in transcripts are relatively subjective and vary across different corpora, thus the unusual initial move structure, including cascading or 'hanging' move structure where "one move is syntactically dependent on the previous one" (Lorés-Sanz, 2016, p.72),

were not considered. Only move chain structures, i.e., move cycles, in EMI lectures of CCL and CEL are examined in the present study.

Through a combination of the *n*-gram function of AntConc 3.5.8 and manual checking, the present study analyzed the move and step codes of EMI lectures in CCL and CEL. Twenty-two and twenty-nine move cycles were identified in CCL and CEL respectively. However, some cycles may appear coincidentally. Those move cycles with *M4* in the non-initial positions, e.g., M5S4-M4, M5S5-M4, are excluded since the moves and steps in the cycle do not belong to the same topic. Besides, some shorter move cycles are embedded in longer ones (e.g., M5S5-M5S5 in M5S5-M5S5-M5S5 and M4-M5S5-M5S5), thus only the longer ones are counted. Consequently, 14 move cycles were identified in CCL and 22 move cycles in CEL. The results are given in Table 4.5.

Table 4.5 Move cycles in CCL and CEL

CCL		CEL	
Move cycle	**Frequency (%)**	**Move cycle**	**Frequency (%)**
M4-M5S4 #	36 (15%)	M4-M5S5 #	36 (11%)
M5S4-M5S4 *	**36 (15%)**	M5S5-M5S7 *	28 (9%)
M4-M5S5 #	31 (13%)	M4-M5S3a #	25 (8%)
M5S4-M5S5 #	19 (9%)	M4-M5S4 #	22 (7%)
M5S5-M5S4 *	**18 (8%)**	M5S3a-M5S5 #	21 (7%)
M5S5-M5S5-M5S5 #	18 (8%)	M5S5-M5S9 *	16 (5%)
M4-M5S3a #	17 (7%)	M5S5-M5S5 *	15 (5%)
M5S3a-M5S5 #	12 (5%)	M5S5-M5S6 *	15 (5%)
M4-M5S5-M5S5 #	10 (4%)	M5S4-M5S5 #	14 (4%)
M4-M6S3 #	9 (4%)	M4-M5S5-M5S5 #	13 (4%)

Table 4.5 Move cycles in CCL and CEL (Cont.)

CCL		CEL	
Move cycle	**Frequency (%)**	**Move cycle**	**Frequency (%)**
M6S3-M5S3a *	**8 (3%)**	M6S1-M5S5 *	13 (4%)
M1-M3S1 *	**7 (3%)**	M7-M5S5 *	12 (4%)
M3S1-M3S2 *	**7 (3%)**	M5S5-M6S1 *	11 (3%)
M5S9-M5S5 #	7 (3%)	M5S9-M5S5 #	11 (3%)
		M4-M6S1 *	10 (3%)
		M5S5-M6S3 *	10 (3%)
		M4-M6S3 #	9 (3%)
		M5S5-M5S5-M5S5 #	9 (3%)
		M6S3-M5S5 *	9 (3%)
		M4-M5S3b *	8 (2%)
		M4-M5S9 *	7 (2%)
		M5S5-M5S5-M5S7 *	7 (2%)

Note. # indicates move cycles shared by both CCL and CEL. * indicates move cycles exclusive to CCL or CEL.

As shown in Table 4.5, 9 of the 14 move cycles identified in CCL also occur in CEL, including *M4-M5S4*, *M4-M5S5*, *M5S4-M5S5*, *M5S5-M5S5-M5S5*, *M4-M5S3a*, *M5S3a-M5S5*, *M4-M5S5-M5S5*, *M4-M6S3*, and *M5S9-M5S5*. They all appear in the Theme Network Building Phase. Among these shared move cycles, seven are two-move cycles and two are three-move cycles. All except one involve moves and steps of *M4* and *M5*. Specifically, *M5S5*, *M5S4*, *M4*, *M5S3a*, and *M6S3* are involved in the 9 shared move cycles. *M4*, *M5S4*, and *M5S5* are among the most frequent moves/steps in these cyclic patterns. In the actual sense, all these move cycles are concerned with lecturing certain topics to students. Simply put, all moves and steps are structured upon elaborating certain topics. Although *M4* is missing in some move cycles, e.g., *M5S4-*

M5S5, *M5S5-M5S5-M5S5*, *M5S3a-M5S5*, *M5S9-M5S5*, there is usually a *M4* nearby,

usually in the positions before the move cycles, except that some other moves and steps

might appear in between, see excerpts 78) and 79).

78) *M5S4-M5S5*

<M4>Ok, and another way we can classify the igneous rocks by the compositions, the mineral compositions.</M4>

<M6S3>Actually we kind of mentioned about this things. Ok that's the way to classify igneous rocks actually based on the magma types, based on which kind of magma they were formed.</M6S3>

<M5S4>…So actually there are three different kind of magmas, granitic, andesitic, and basaltic, basaltic magmas. And the different magmas have different kind of compositions…So the magma compositions can determines the characteristics, the physical or chemical properties of the magma self and the igneous rocks, solidified from the magma, from magmas…</M5S4>

<M5S5>First the basaltic igneous rocks. Normally the basaltic igneous rocks has a higher percentage of iron and magnesium. And in that case normally they have a kind of dark color, dark color, and high density, dark color and high density because these items are actually heavier than the silica, right? Than silica…And it is fluid and flows freely from volcanoes in Hawaii so we can actually finds lots of basaltic igneous rocks in the, in Hawaii. And the basalt is the most common rock type in the earth's crust, in the earth's

crust...Ok the basalt, basaltic rocks. They have the high composition of iron

and magnesium, but less silica, less silica. And the color is normally dark. And

the density is dense. And we can find lots of them from the ocean floor, from

the ocean floor, and from Hawaii, and from Hawaii. And also we can actually

find many of them from the Africa, the Great Africa rist, Rift sorry, Great Africa

Rift. </M5S5> (CCL02)

In excerpt 78), the lecturer initiated the topic of classifying the igneous rocks by

the compositions in *M4*, then she connected the present topic with something mentioned

before in the lecture in *M6S3*. After that, she continued her elaboration of the

mechanism of magma properties in *M5S4*. And she further provided factual knowledge

of one typical igneous rocks, i.e., the basaltic the basaltic igneous rocks, in *M5S5*. Thus

the move cycle of *M5S4-M5S5* may function jointly with other moves and steps in the

elaboration of a topic.

79) *M5S4-M5S5*

<M4>okay... we now need to select amongst the variation [3.0]</M4>

<M5S3a>again [0.5] on first thought that seems [0.2] kind of easy [0.3]...but

first of all you have to realize that many characters are actually quite difficult

to measure...the second [0.3] er [1.4] thing [0.4] you need to bear in mind is

that you've got a lot of material to look at [0.4] a lot of different ones that you

want to assess...and the other problem is that you usually have small quantities

of it [1.0]</M5S3a>

<M5S4>okay [0.2]…you can't totally so you have to pick a single condition [0.5]…you've only got a small amount of material [1.8] a small amount of material a… so [0.3] the relevance [0.2] of the characters as well [0.3] as the effectiveness that you can measure them [0.2] is extremely important [4.5]</M5S4>

<M6S3>the next problem we have [0.8] is really the one that we were talking about a few minutes ago…the efficiency with which we can select…</M6S3>

<M5S4>as i think is fairly obvious from th-, [0.2] from the equation [0.5] the greater the environmental effects [0.4] the more the phenotype [0.8] will be a poor reflection of the underlying genotype [1.1] okay [0.4] so the more something's affected by the environment [0.2] the less we can do anything about [0.5] er [0.2] picking it …</M5S4>

<M5S5>so if you think for instance [0.2] the character that's similar in humans is height [0.8] you are all different heights [0.9] but [0.3] the fact there's not one of you here that's a major g-, has a major gene difference for height in other words there's no achondroplasic dwarves [1.0] but [0.2] they do occur [0.3] in the population there are major genes for height [0.4] but most of you differ [0.3] for height [0.6] by a lot of genes [0.7] there's a lot of different genes will affect your height [0.8]</M5S5> (CEL05)

In excerpt 79), the lecturer started the topic of selecting variation in plant breeding in **M4**, then in **M5S3a**, he/she stated the background of doing the selection, which is

deemed difficult. After that, the lecturer went on to talk about the mechanism of the difficult selection in *M5S4*, which is connected through *M6S3* with the issue of selection efficiency that was mentioned a few minutes before in the lecture. It is only after these moves and steps that the lecturer began to talk about the issue of variation selection in *M5S4*, where he/she expounded on effect of the environment on the selection. And the lecturer further explained with analogies of human height to show the difficulty in variation selection of plant breeding. This move cycle of *M5S4-M5S5* work jointly with other moves and steps to corroborate the topic of variation selection in *M4*.

More importantly, Table 4.5 also reveals exclusive move cycles in CCL. Five move cycles were found unique to CCL lecturers, including *M5S4-M5S4*, *M5S5-M5S4*, *M6S3-M5S3a*, *M1-M3S1*, and *M3S1-M3S2*. Two of them (*M1-M3S1*, and *M3S1-M3S2*) are in the Opening Phase, whereas three (*M5S4-M5S4*, *M5S5-M5S4*, and *M6S3-M5S3a*) occur in Theme Network Building Phase.

Except 80) demonstrates how the move cycle *M1-M3S1* is employed by the Chinese EMI lecturer.

80) ***M1-M3S1***

<M1>S1: Ok now class begin. Good afternoon!</M1>

<M3S1>Today today we are going to study a new concept. Oh that is er infinite sequence and its limit, Are you familiar with such a notion?

SS: Yes yes

S1: Yeah, good. yeah.</M3S1> (CCL12)

In excerpt 80), the Chinese lecturer announced the official start of the lecture through **M1**, but he didn't include the conventional **M2 Warming Up**, instead went on directly to **M3S1** to announce the theme of the lecture. This non-linear move structure is considered one typical characteristics of spoken data in ELF contexts.

81) **M3S1-M3S2**

<M3S1>Then today we are going to start a new chapter. This chapter three, basically Advanced Hydraulic Fracturing Technology.</M3S1>

<M3S2>So here is my outline, so for this seven sessions, today we are going to talk about the review of hydraulic fracture. Then we are going to, probably if we have time, we are going to talk about multi-layer fracturing, then tips screen out technology, fracture height control. But today we are going to focus on the first one, review of hydraulic fracture.</M3S2> (CCL08)

The move cycle **M3S1-M3S2** in excerpt 81) is, to a large extent, the result of the conventional status of these two individual steps in CCL. **M3S1** and **M3S2** appear in 92% and 75% of CCL lectures respectively, whereas they only occur in 67% and 33% of CEL lectures. The emergence of new moves/steps and move cycles, in this case **M3S1-M3S2**, is another feature of ELF contexts.

82) **M5S5-M5S4**

<M5S5>For example, Civil War happened between eighteen sixty one to eighteen sixty five, and World War two happened during nineteen thirty nine to

nineteen forty five, so it's absolute date, and we also can make the right order,

the Civil War happened before the the World War two. The same idea about the

Shihezi Formation.</M5S5>

<M5S4>That's different way to relative time about geology. So when we talk

about the characteristic, chronological time, we pronounce another terms,

chronological dating, chronological dating, or simply dating, is the process of

attributing to an object or event a date in the past, so those event will be located

in a previously established chronology. En we just write the whole chronology,

er, in this region. And then we can put different events into this

chronology.</M5S4> (CCL01)

In excerpt 82), the Chinese lecturer employed the **M5S5-M5S4** move cycle to help explain the topic of geological time scale. She first used examples to illustrate the topic through **M5S5**, and continues to explain the mechanism of the topic through **M5S4.** The move cycle **M5S5-M5S4** is rarely used by native English lecturers in CEL, thus could be another feature of move patterns in ELF contexts.

83) **M5S4-M5S4**

<M5S4>Ok so first the as we can see here, the spreading center…And in that

case, the pressure is lower…when the pressure is decreased, decreased, sorry,

the magma is formed…That's the area how the magma can raised up and the

igneous rocks can formed in this area. So first in the spreading center. That's

the area how the magma can be formed and when the temperature going to be

lower or the pressure is higher than, it can be lithificated as igneous rocks. Ok

that's the first environment.</M5S4>

<M5S4>The second environment, hot spot, hot spot…And in that case, this

area the magmas is formed. And when the magmas is formed…And then later

the igneous rocks can be lithificated from the magma.</M5S4> (CCL02)

In excerpt 83), the Chinese lecturer referred to the move cycle **M5S4-M5S4** to

expound the mechanisms of how magmas could be converted to igneous rocks. This

move cycle is not found in CEL, thus could be very likely the new move pattern in ELF

contexts.

84) ***M6S3-M5S3a***

(And let's study about …choke performance.)

<M6S3>So we have learned about the IPR, we have learned about TPR, and

number three is choke performance, ok.</M6S3>

<M5S3a>So we can see this device is a choke, ok, er, this device is a choke.

The choke is just we fixed on the wellhead, ok, to control the, control the flow

rate, control the production rate, ok. So here, er, this, we call it Christmas tree,

right? Er, Christmas tree. Er, so this is a schedule of the wellbore, you can see

here the yellow channels is perforation, ok, er, perforation. The perforation

connect the wellbore and the reservoir, right? So the fluid can flow through the

channels into the wellbore, ok, er. And this green is casing, er, we have three

types of casing. Surface casing, intermedia casing, and production casing, ok.

And this yellow part is what? That's tubing, ok. Er, I I have said the oil flow

through the tubing to the surface, ok, er tubing. And here we have a packer, ok.

Isolate the, er space in the annulus, ok. So the packer separate the annulus

space into two part, right, into two part? Ok, er.</M5S3a> (CCL07)

In excerpt 84), the Chinese lecturer employed the move cycle **M6S3-M5S3a** to help elaborate the topic of choke performance. The lecturer first connected choke performance with other topics such as IPR and TPR through **M6S3**, then through **M5S3a** he continued to provide background information of the device 'choke' as well as other related wellbore devices such as 'perforation', 'casing' and 'packer'. These devices work together to control the well production rate. This new move cycle never occurred in CEL, thus is another piece of evidence of move cyclicity in ELF contexts.

4.6 Summary

This chapter reported the results and discussion of move analysis conducted on EMI lectures in CCL and CEL. Twelve moves with individual steps were identified in the two corpora. An overall log likely-hood comparison showed no significant difference (G2= + 3.58, p>0.05) in terms of the number of moves and steps in lectures of CCL (685) and CEL (1161). In fact, log likely-hood result suggests slightly more moves and steps used by Chinese lecturers than by English lecturers, indicating that they are aware of the moves and steps in lecture genres, at least, as well as their native English counterparts. In addition, interview data suggested genre transfer of Chinese

EMI lectures from their Chinese lectures.

All moves/steps (except that there is no *M2S3* in CCL and *M2S1* CEL) appear in both CCL and CEL. Most moves/steps in the two corpora share similar conventionality status, though there are some differences throughout the lectures. It seems to suggest that CCL and CEL share similar rhetorical move structure.

However, different lecturing preferences can still be found in the use of individual moves and steps between CCL and CEL lecturers, with the major ones centering upon the Theme Network Building Phase.

There are three moves in the Opening Phase: *M1 Getting Started*, *M2 Warming Up*, and *M3 Setting Up Lecture Agenda*. Chinese lecturers' conventional use of *M1* seems to be indication of their inclination to practices of native English language lecturers in this move. In addition, Chinese EMI lecturers used significantly more *M2S1 Leading in* and *M3S2 Providing lecture scope*, which could be influenced by and transferred from the common practices while they give lectures in Chinese.

As the major part of academic lectures, the Theme Network Building Phase consists of six moves: *M4 Introducing The Topic*, *M5 Elaborating The Topic*, *M6 Building Theme Network*, *M7 Making Aside*, *M8 Housekeeping*, and *M9 Checking Comprehension & Consolidating Learning*.

Influenced by course level, students' level, personal lecturing style and the topic-orientation of EMI courses, Chinese EMI lecturers used significantly more *M4* than the native English lecturers.

The ten steps of *M5 Elaborating the topic* exhibit different preferences between Chinese and English lecturers. Generally, Chinese lecturers tend to rely more on text books, prepared power-point slides as well as authoritative resources, which lead to significantly more use of *M5S1 Explaining term*s and *M5S4 Expounding rationale*. In addition, Chinese lecturers' over-reliance on single type of background-related sub-step lead to significantly more use of *M5S3a Presenting background knowledge*. On the other hand, perhaps due to a wider repertoire of background-related sub-steps, native English lecturers show significantly greater use of *M5S3b Contextualizing real-word experience*. English lecturers also use significantly more *M5S6 Providing caveats*, which might be due to Chinese EMI lecturers' incapability of doing so though interview data showed unanimous agreement on the necessity of this step among the interviewees. Meanwhile, English lecturers' much greater use of *M5S7 Making comments* could be related to cultural differences between Chinese and English lecturers. In addition, English lecturers also employ significantly more *M5S10 Pinning down and/or clarifying the topic*, which could be related to the emerging status of EMI courses in China, whose lecturers are less experienced in offering clarifications to help students.

The fact that the twelve CCL lectures were all co-taught by different lecturers might have undermined the basis of *M6S1 Referring to previous lecture (part)/other Source*, thus lead to much less *M6S1* among Chinese EMI lecturers.

Different attitudes towards textbooks and different classroom practices between

Chinese and English lecturers may have contributed to much more use of *M8 Housekeeping* by native English lecturers.

Different strategies of comprehension checking may have led to significantly less use of *M9 Checking Comprehension & Consolidating Learning* among Chinese EMI lecturers. All other moves and steps in this phase didn't show significant difference among lecturers in the two corpora.

The Closing Phase consists of three moves: *M10 Wrapping Up Lecture*, *M11 Cooling Down*, and *M12 Ending Lecture*. The only difference between CCL and CEL lies in the use of *M10*. There is significantly more *M10* found in CCL than in CEL, which might be due to Chinese lecturers' preference to provide students with definite conclusions out of their own inductive thinking.

Different from the long-term evolving process in academic writing, an ELF perspective of CCL lectures didn't find any move simplification among Chinese EMI lecturers. However, move hybridity in the form of move cycles did occur in CCL, where five move cycles were found unique to CCL lecturers, including *M5S4-M5S4*, *M5S5-M5S4*, *M6S3-M5S3a*, *M1-M3S1*, and *M3S1-M3S2*.

CHAPTER 5

RESULTS AND DISCUSSION II: COMPARISON OF

FORMULAIC SEQUENCES IN CCL AND CEL

This chapter presents the results and discussion of the investigation of formulaic sequences used in CCL and CEL lectures. First, it provides an overview of formulaic sequences used in CCL and CEL lectures. Then, it presents the most frequent formulaic sequences in the two corpora. The third part of this chapter demonstrates the structural distributions of formulaic sequences. The following part examines formulaic sequences from an ELF perspective, and the last part ends the chapter with a summary of results and findings of formulaicity analysis in CCL and CEL lectures.

5.1 Overview of formulaic sequences in CCL and CEL

The results and discussion on formulaic language use in CCL and CEL lectures are expected to answer research question 2), which is presented again as follows:

2) What are the similarities and differences in the use of formulaic sequences in university EMI lectures given by native and non-native English lecturers?

Table 5.1 Statistics of FSs in CCL and CEL lectures

Lecture	word count	FS tokens	FS types	FS frequency (pmw)	Lecture	word count	FS tokens	FS types	FS frequency (pmw)
CCL01	3785	191	116	3377	CEL01	7496	267	191	2531
CCL02	4606	191	67	3377	CEL02	6751	338	202	3204
CCL03	4109	100	62	1768	CEL03	7048	283	194	2682
CCL04	4735	106	50	1874	CEL04	5924	271	155	2569
CCL05	4524	170	72	3006	CEL05	9616	418	230	3962
CCL06	6807	198	73	3501	CEL06	6276	239	140	2265
CCL07	4911	92	53	1627	CEL07	15713	675	381	6398
CCL08	7073	202	60	3572	CEL08	9708	503	276	4768
CCL09	4552	113	50	1998	CEL09	12212	479	242	4540
CCL10	3917	56	32	990	CEL10	10939	401	193	3801
CCL11	4473	129	41	2281	CEL11	5312	178	134	1687
CCL12	3064	67	43	1185	CEL12	8057	302	175	2863
Total	56556	1615	546	28556	Total	105052	4354	1688	41269

Table 5.1 illustrates the overall statistics of formulaic sequences identified in EMI lectures of CCL and CEL. Specifically, the word count of each lecture, the formulaic sequence types and tokens, and the per million words (pmw) frequency were given. A total of 1615 and 4354 formulaic sequences were identified in CCL and CEL lectures, respectively. On average, the Chinese EMI lecturers used 28566 FSs per million words whereas the native English lecturers used 41269 FSs per million words. The log likely-hood comparisons of both types (G2= -115.97, p< 0.0001) and tokens (G2= -172.37, p< 0.0001) indicate significantly less use of FSs by the Chinese EMI lecturers than their native English counterparts. Chinese EMI lecturers seem to use comparatively much less formulaic language while lecturing. A tentative speculation might be that

Chinese EMI lecturers possess a more limited repertoire of FSs than the native English lecturers. Interview data corroborated such speculation, with all four interviewees admitting limited FS repertoire in comparison with native English lecturers. As interviewee A put it, "we have limited vocabulary", and "the small number of phrases we possess should meet the instructional needs of the courses, so we mostly rely on a limited repertoire of these common phrases. Even when we use these phrases, we do not use them as freely as the native English lecturers". Though no previous study has compared FS use between native and non-native English speakers in spoken genres or registers, much research on written data has already revealed this discrepancy between them (e.g., Chen & Baker, 2010; Ädel & Erman, 2012).

5.2 The most frequent formulaic sequences in CCL and CEL

Though frequency is not the yardstick in the identification of formulaic language, the present study lists two tables of top 20 most frequent formulaic sequences, so as to give a quick glimpse of the formulaic sequences used in the two corpora. Meanwhile, a dispersion threshold of occurring in 2 out of 12 lectures was adopted to avoid individual lecturer/discipline idiosyncrasies.

Table 5.2 Top 20 most frequent FSs in CCL and CEL

Rank	CCL rank	FS	Freq.	Lectures	CEL rank	FS	Freq.	Lectures
		CCL				**CEL**		
1	1	*we can see*	79	8	1	***there are***	97	12
2	2	***you can see***	67	5	2	*there's*	92	11
3	3	*that means*	60	6	3	***have to***	89	11
4	4	***for example***	60	9	4	***so that***	69	11
5	6	***there is***	34	8	5	*of course*	67	11
6	7	*let's see*	33	3	6	***I think***	62	11
7	8	***there are***	27	6	7	***a lot of***	58	11
8	9	***look at***	27	2	8	***look at***	55	10
9	10	***so that***	26	2	9	*in terms of*	52	10
10	11	*let's look at*	20	4	10	***for example***	45	9
11	12	***you know***	18	6	11	*in fact*	40	11
12	13	*that's why*	17	3	12	*I mean*	39	8
13	14	*in that case*	16	2	13	***there is***	35	8
14	15	***a lot of***	16	4	14	*equal to*	35	2
15	17	***I think***	14	5	15	*in other words*	33	7
16	18	*this kind of*	13	5	16	***you know***	32	7
17	19	*because ... so*	13	6	17	***you can see***	31	8
18	20	*the definition of*	12	2	18	*and so on*	29	6
19	21	***have to***	11	5	19	*kind of*	28	7
20	23	*according to*	11	5	20	*I'm gonna*	25	3
					21	*as well*	25	9

Table 5.2 lists the top 20 most frequent FSs in CCL and CEL. As revealed in CCL and CEL rank column of Table 5.2, some FSs were eliminated from the top 20 most frequent FS list. Those excluded FSs are legitimate formulaic sequences since they meet the criteria stipulated in Section 3.4.1, but they were excluded from the list because the focus here is on the most frequent FSs and individual lecturer/discipline idiosyncrasies

need to be sifted out. For example, *by the wind* in CCL was not selected in the list because it occurred only in CCL04: *Introduction to Physical Geology-Sediment Transport and Deposition*, despite a high frequency of 11 occurrences. This sequence is very closely related to the discipline of geology, when the lecturer used it to explain the formation of sedimentary rocks.

It was shown that 10 of the top 20 most frequent FSs were shared (italicized bold lettered FSs) by the two groups of lecturers. They include: *you can see*, *for example*, *there is*, *there are*, *look at*, *so that*, *you know*, *a lot of*, *I think*, and *have to*. Most of these shared FSs are either verb-related phrases or clausal phrase fragments, which are characteristic of classroom teaching (Biber et al., 2004). In terms of length, they are all 2- or 3-word short FSs, a typical feature of spoken language.

The prepositional code gloss *for example* is used to supply additional information to ensure students' comprehension, which is characteristic of classroom discourse. In addition, the use of the adjective phrase *a lot of* instead of *many* is also indicative of typical features of oral language.

The more illuminating aspect of FSs lies in the examination of those special FSs that are exclusive to either CCL or CEL lecturers. Table 5.3 presents the top 20 most frequent FSs exclusive to CCL and CEL.

Table 5.3 Top 20 most frequent FSs exclusive to CCL and CEL

Rank	CCL				CEL			
	CCL rank	FS	Freq.	Lectures	CEL rank	FS	Freq.	Lectures
1	7	*Let's see*	33	3	2	*there's*	92	11
2	11	*Let's look at*	20	4	14	*equal to*	35	2
3	13	*That's why*	17	3	15	*in other words*	33	7
4	19	*because ... so*	13	6	20	*I'm gonna*	25	3
5	24	*We're going to talk about*	10	2	21	*as well*	25	9
6	30	*different kind of*	9	3	22	*what happens*	22	6
7	36	*we see*	7	3	23	*we're going to*	21	8
8	43	*we will talk about*	6	4	25	*we're gonna*	19	4
9	48	*I show you*	6	2	26	*at least*	19	7
10	53	*we will see*	5	3	27	*go through*	18	6
11	56	*three kind of*	5	3	28	*as I said*	18	8
12	54	*we are going to talk about*	5	2	36	*tend to*	15	7
13	59	*if we want to*	5	2	37	*looking at*	15	6
14	61	*figure out*	5	2	38	*I'm not going to*	15	5
15	80	*belong to*	4	3	39	*I'm going to*	15	5
16	67	*We'll talk about*	4	2	42	*talk about*	14	7
17	74	*is depends on*	4	2	43	*work out*	13	4
18	82	*another kind of*	4	2	44	*there were*	13	4
19	96	*not only ... but also*	3	3	45	*the fact that*	13	8
20	85	*we talk about*	3	2	47	*is going to be*	13	5

A notable difference between the two groups of lecturers lies in the use of FSs with personal pronouns 'we' (hereinafter referred to as 'we' FS) and 'I' (hereinafter referred to as 'I' FS).

As manifested in Table 5.3, ten types of FSs containing either the subjective or

objective forms of 'we' were found in the CCL list, whereas only two types of 'we' FSs were in the CEL list. The cumulative frequency of these 'we' FSs has reached 98 and 40 occurrences for CCL lectures and CEL lectures, respectively. The log likely-hood comparison (G2= +74.09, p< 0001) confirms significantly more use of 'we' FSs by the Chinese EMI lecturers than their native English counterparts. Structurally, these 'we' FSs take three different patterns, i.e., '**Let's + VP**', '**we + VP**', and '**if + we + VP**', which are all earmarks of oral speeches. Functionally, these FSs are used to state aims and objectives 85), introduce topics 86), engage students 87), and explain concepts 88).

85) *And next class **we will talk about** another empirical method, that is the decline, decline analysis*. (CCL09)

86) *Here **we will see** the application of the waterflooding curve*. (CCL09)

87) *We, **if we want to** know productivity, we must do the well monitoring*. (CCL10)

88) *And **let's see**, the liquid droplet move, drop very fast, because, you know, the gas, if the gas flow is not fast, or the flow rate is not big. You see this, the droplet may flow, fall down very slowly because of, you know, this increase this one. If the gas flows faster, what will happen, the droplet may stop there, right? And what he said, ok, if because the droplet will stop falling, that's minimum gas velocity we need, right? That means for your gas well, if you have higher gas flow rate, or gas velocity, then all this liquid will move to the well*. (CCL06)

Although the 'we' in 'we' FSs may have various referents (e.g., 'inclusive we',

'exclusive we, 'we' for 'you'), 'we' FSs are generally used to reduce the distance in the lecturer–student relationship, achieve solidarity and to bring them together as a single community (Yeo & Ting, 2014). In particular, engagement markers in the '**Let's + VP**' construction serve to include students as discourse participants in the classroom community, which may greatly facilitate harmonious teacher-student relationship.

On the other hand, the use of 'I' FSs may just have the opposite effect. Self mention marker 'I' in 'I' FSs projects an authoritative identity of lecturers on students, which tends to draw a separating line between lecturers and students. Although there are only six occurrences of 'I' FS use (*I show you*) in CCL, a total of seventy-three occurrences of 'I' FS (*I'm gonna*, *as I said*, *I'm not going to*, and *I'm going to*) were extracted from CEL. The log likely-hood comparison (G2= -33.02, p< 0001) indicates significantly less use of 'I' FSs by the Chinese EMI lecturers than their native English counterparts.

In terms of functions, *I show you* is generally used to signpost the process of the lecture 89), or establish the lecturers' authority 90), or sometimes both.

89) *So **I show you** the number, most of the US gas well, they are using three, two three inch tubing so this, actually this tubing is the most popular used in the United States.* (CCL06)

90) *Ok, er, the slide, slides **I show you** the different kinds of sedimentary and bedding, and the sedimentary usually transport media is water, air, ice, and gravity.* (CCL004)

In except 89), the lecturer explicitly tells students the actual process of the lecture,

informing students of things he/she intends to do. Excerpt 90) starts the sentence with some hesitation fillers and repeats '*Ok, er, the slide, slides*', then the lecturer self-corrected '*slide, slides*' into '*I show you*' to highlight the authority of the lecturer as the source of the information.

The 'I' FSs in CEL are a bit more complicated since they serve different functions. *I'm going to*, together with its contracted form *I'm gonna*, usually serves to signpost the process of the lecture 91), give instructions 92), and make announcements 93).

91) *right [0.3] now [0.2]* **I'm going to** *[0.4] specialize now and talk particularly about kidney transplantation because that's the subject of these lectures.* (CEL07)

92) *now* **I'm going to** *make my big assumption [1.2]* **I'm going to** *assume [0.9] that [1.1] this thing rolls [1.7] and doesn't slip [1.7] and I've [0.2] put a note there that i must check [0.5] at the end [0.5] that that is a reasonable assumption to make [2.3] or at least check for consistency [1.4] we'll see how [0.3] we do that when we get there [4.0]* (CEL02)

93) *and next Thursday I'm also gonna assign I mean this Thursday next lecture* **I'm gonna** *assign uh practice problems for homework.* (CEL10)

However, the negative form *I'm not going* is commonly used to pin down and/or clarify the topic 94-95). It is often followed by verb phrases denoting unnecessary instruction or discussion, e.g., 'worry too much', 'talk too much about', 'go through it in great detail', 'talk about that in detail'.

94) ***I'm not going to*** go through it in *[0.6] great detail [0.4] but i want you to [0.4]*

focus your attention [1.4] on this column here [2.6] (CEL06)

95) *and [0.3]* ***I'm not going to*** *[0.3] labour the technical side of radar [0.4] the*

[0.9] Met students will get it i think in their third year from [0.4] Dr namex.

(CEL04)

In addition, the 'I' FS *as I said* in the CEL list is frequently used to activate students'

prior knowledge 96) or establish common ground for the instruction 97).

96) *okay [2.4] now unfortunately* ***as I said*** *last week [1.0] having the modulus [1.0]*

in the [2.7] fibre direction [1.8] one [1.4] and having the modulus in the

transverse direction [0.2] two [0.8] is not enough [1.0] to get all the

information we need [0.6] to do any calculation [0.6] relating to design stress

analyses or whatever [0.2] of [0.8] er [0.5] reinforced [0.3] plastic [0.8]

materials [1.2] we need [0.2] (CEL03)

97) *the reason I've laboured in going through this is because* ***as I said*** *to you [0.3]*

when you go out of here with a degree in microbiology and virology [0.3] and

you tell somebody you know something about viruses [0.6] almost certainly

the virus that they will ask you about is H-I-V [0.7] and they will expect you

to be reasonably informed and what they won't want to know from you [0.3]

is how tat and rev works [0.6] (CEL06)

The last facet to note is the non-native like FSs *because…so, three kind of* and *is*

depends used by Chinese lecturers. Though these three FSs take non-standard lexico-

grammatical forms, they are all semantically transparent and straightforward, thus do not seem to cause any confusion or disturbance in communication. All four interviewees firmly articulated that such non-standard use of FSs does not affect their classroom communication or instruction at all, though interviewee C contended that more standard FS use is preferable for university lecturers since it affects students' judgment of and trust on teachers.

In essence, these non-native like FSs all belong to the approximated FSs that lead to no disturbance in Björkman (2008b). Functionally, they serve exactly the same functions as their standard counterparts in the context, as suggested in excerpts 98-100).

98) *Basically it's the volume contac- contact **because** your hydraulic fracture, **so** stimulated volume, we can know it through the micro seismic, so that's basically a good thing*. (CCL08)

99) *Er, there are **three kind of** rock types in our nature. The igneous rocks, sedimentary rocks and metamorphic rocks*. (CCL05)

100) *And the intercept of the A curve **is depends on** the geological reserve and the viscosity ratio between oil and water*. (CCL09)

5.3 Structural categories of formulaic sequences

The structural categories of formulaic sequences are given in Table 5.4. It reveals that proportionally EMI lecturers in CCL and CEL didn't show much difference in most structural categories (except that Chinese lecturers used proportionally more clausal

FSs but less VP-based FSs than native English lecturers), suggesting high congruency in terms of structural distribution of FS use between the two groups of lecturers. In other words, Chinese EMI lecturers may have similar allocation of FS structure categories to native English lecturers.

However, the actual occurrences (token frequencies) of the different structural categories show significant differences between CCL and CEL. Tables 5.4 indicates that Chinese EMI lecturers used significantly less NP-, PP-, AP(C)-, VP-, and AdP-based FSs, implying relatively limited FS repertoire in these structural categories among the Chinese EMI lecturers.

Table 5.4 Structural distribution of formulaic sequences in CCL and CEL

Structure category	Types		Tokens		Log Likelihood Critical Value G2
	CCL (%)	CEL(%)	CCL(pmw)/%	CEL(pmw)/%	
NP	58(11%)	177(10%)	115 (2033)/7%	327 (3099)/8%	-16.43****
PP	131(24%)	402(24%)	372 (6578)/23%	1181 (11194)/27%	-88.52****
AP(M)	1(0%)	2(0%)	3 (53)/0%	3 (28)/0%	+0.57
AP(C)	19(3%)	58(3%)	36 (637)/2%	141 (1336)/3%	-18.27****
Clausal	217(40%)	561(33%)	802 (14181)/50%	1522 (14426)/35%	-0.24
VP	100(18%)	441(26%)	209 (3695)/13%	942 (8929)/22%	-159.7****
AdP	7(1%)	25(1%)	21 (371)/1%	139 (938)/3%	-39.44****
ConjP	13(2%)	19(1%)	57 (1008)/4%	99 (938)/2%	+0.16
Total	546(100%)	1685(100%)	1615 (28556)/100%	4354 (41269)/100%	-172.37****

Note. $^*p < 0.05$, critical value G2= 3.84; $^{**}p < 0.01$, critical value G2 = 6.63; $^{***}p < 0.001$, critical value G2 = 10.83; $^{****}p < 0.0001$, critical value G2 = 15.13.
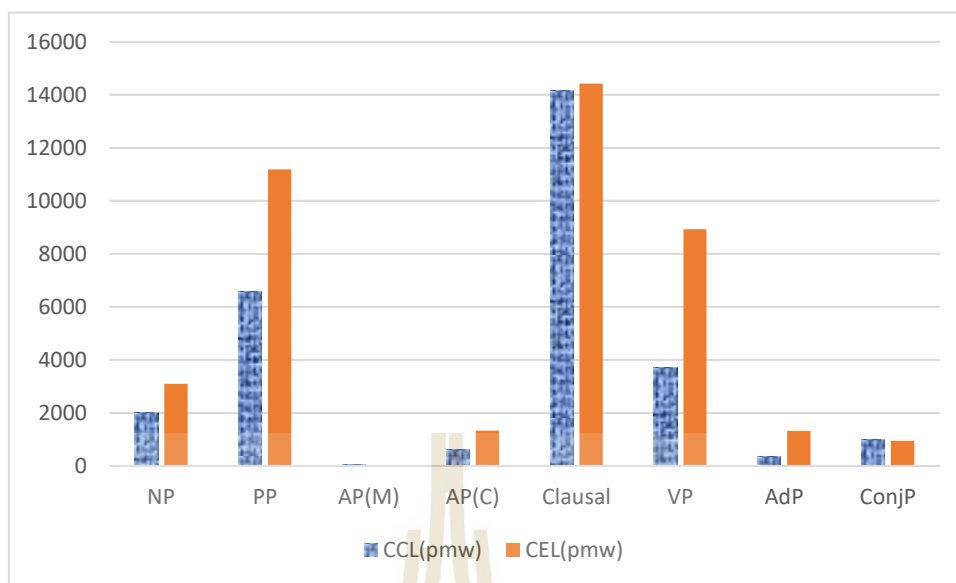
Figure 5.1 Structural distribution of FSs (pmw) in CCL and CEL

Figure 5.1 visually displays the major structural categories of FSs used by native Chinese and native English EMI lecturers. In term of token frequencies, lecturers mainly relied on NP-, PP-, clausal, VP-, and ConjP-based FSs, which will be discussed individually in detail.
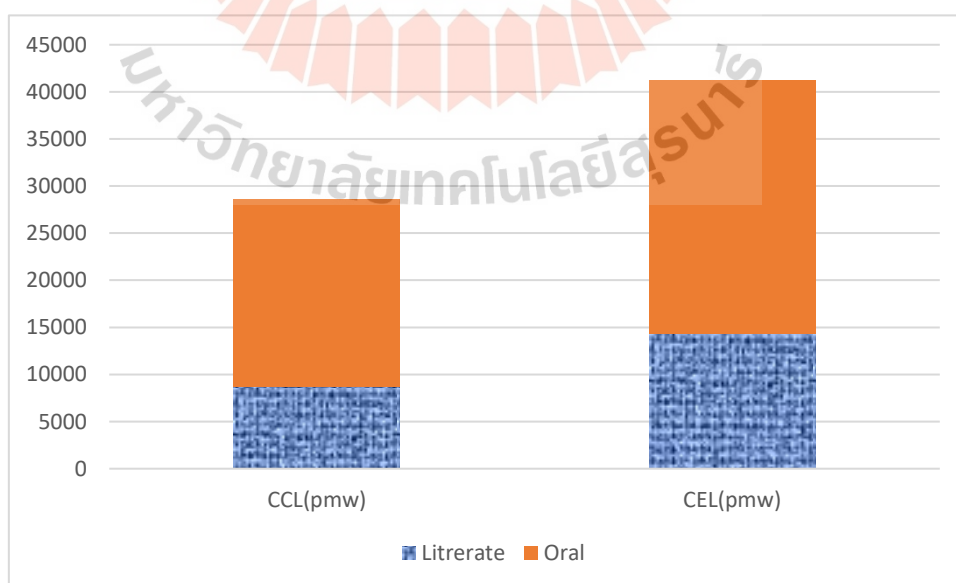
Figure 5.2 Lecturing preferences in CCL and CEL

According to Biber et al. (2004), classroom teaching exhibited a mixture of 'oral' (VP and clausal bundle use) and 'literate' characteristics (NP and PP bundle use). This is also attested in the present study. Although other structural categories of FSs take only very small proportions, they were also taken into consideration when the researcher analyzed the overall lecturing preferences. The use of NP-, PP-, and AP(M)-based FSs were considered 'literate' features of lectures, whereas the VP-, clausal, ConjP-, AdP-, and AP(C)-based FSs the 'oral' features. Figure 5.2 shows that both CCL and CEL exhibit a mixed feature of 'oral' and 'literate' characteristics.

5.3.1 Noun phrase fragment formulaic sequences

As shown in Table 5.4, Chinese EMI lecturers used approximately the same proportion of NP-based FS types as the native English lecturers, but significantly less tokens than the latter (G2=-16.43, p<0.0001). Table 5.5 lists the top 20 most frequent NP-based FSs in the two corpora. The dispersion threshold of occurring in at least two lectures is still adopted here as well as in the discussion of other structural categories.

Table 5.5 Top 20 most frequent NP-based FSs in CCL and CEL

Rank	CCL			CEL		
	FS	Freq.	Lectures	FS	freq.	Lectures
1	*the definition of*	12	2	*the amount of*	15	5
2	*each other*	5	2	*the number of*	14	4
3	*the bottom of*	4	3	*the fact that*	13	8
4	*the type of*	3	2	*the value of*	9	4
5	*the direction of*	3	2	*the range of*	9	4
6	*the end of*	2	2	*each other*	9	4
7	*the difference between*	2	2	*the idea of*	7	3
8	*the amount of*	2	2	*the end of*	6	4

Table 5.5 Top 20 most frequent NP-based FSs in CCL and CEL (Cont.)

Rank	CCL			CEL		
	FS	Freq.	Lectures	FS	freq.	Lectures
9				*the probability that*	6	2
10				*the use of*	5	3
11				*the shape of*	4	3
12				*the effect of*	4	3
13				*difference between*	4	3
14				*the sort of*	4	2
15				*the size of*	4	2
16				*one of the things*	3	3
17				*things like that*	3	2
18				*the thing to notice*	2	2
19				*the sum of*	2	2
20				*the strength of*	2	2
				the sorts of	2	2
				the principle of	2	2
				the definition of	2	2
				the bottom of	2	2
				something like this	2	2

Structurally, most of the NP-based FSs are in the pattern of **'*the* + *noun* + *of*'**. Most of the FSs are used to describe the physical properties of objects 101), the concepts 102) and the procedures 103) of experiments or projects. It seems that the native English lecturers used much more diverse FSs than the Chinese lecturers, therefore might be better able to explain properties, concepts and procedures involved in scientific experiments and industrial projects.

101) *The term can refer to **the amount of** ore remaining in a particular mine,*

or it can be used on a global or national scale. (CCL03)

102) *so the **the idea of** the grating is to [0.4] er not really [0.2] give much [1.0] er [0.2] insplash but it [0.2] gives you a a more smooth aerodynamic surface so you don't get too much eddying [0.6]* (CEL04)

103) *er we've seen a [1.4] a massive [1.4] growth in **the use of** radars for rainfall growth in [1.3] rainfall radars [2.7] and particularly in developed countries [0.5]* (CEL04)

A second observation is the use of vague expressions by the native English lecturers, e.g., *things like that*, *something like this*. Concordance analysis reveals that these vague expressions are generally used to succinctly refer to something that has just been mentioned. Such expressions could be conducive to increasing the efficiency of classroom teaching given the limited lecture time spans. However, this type of FSs are generally lacking among the Chinese lecturers.

104) *but [0.2] to bring my [0.3] me to the s-, final conclusion [0.8] er [2.4] plainly [0.3] spare part surgery [0.3] and bone marrow transplantation and **things like that** [0.3] does work [0.2] whereas [0.6] technically a-, [0.2] and ethically possible it does work [1.1]* (CEL07)

105) *now once we've done that we've done the swap, and actually let's use the, board here for a minute. <PAUSE WHILE WRITING ON BOARD> kay it's pretty clear we're doing, **something like this**, to trade the two elements* (CEL 12)

The next thing to notice is the use of FSs that incorporate shell nouns, which

can perform characterization, temporary concept-formation, and linking functions (Schmid, 2000; Arktas & Cortes, 2005). Chinese lecturers didn't use any of this type of FSs, whereas native English lecturers used the shell noun FSs to characterize the experience or process of an experiment 106), to form a temporary concept 107) or link contrasting or complex ideas 108). All these are, in essence, to increase the coherence of ideas or concepts in lecturing.

106) *right well yesterday i was er [0.5] talking about **the [2.2] idea of** a track [0.3] in [0.2] radiation chemistry [0.4] where [0.3] as the particle moves along [0.6] it's losing energy* (CEL01)

107) *in fact, we measure, **the effect of** metabolic rate, on temperature uh temper- we we- measure, quantify **the effect of** temperature on metabolic rate by a quantity that we call, Q-ten, which is the rate at any temperature... at uh ten degrees higher, than that at a, lower temperature.* (CEL08)

108) *the other one [0.5] is that [0.6] in this calculation one neglects **the fact that** a Poisson's ratios [0.5] all the fibres in the matrix are [0.7] different [0.6] when you are pulling in that direction [0.5]* (CEL03)

The last point to mention is the use of FSs that stress topic importance 109-110). The native English lecturers' use of this type of FSs demonstrates their richer repertoire of formulaic language in stressing the importance of classroom contents.

109) *okay so the **the thing to notice** then is that this is this is no longer a*

straight, age class model, because now we're lumping individuals of

several ages together in this final category, uh cup- an adult stage.

(CEL11)

110) *so [0.2] clearly **one of the things** that that [0.3] people have looked for*

[0.6] in [1.4] trying to study this infection [0.3] is prognostic indicators

[0.3] of the length of period [0.7] of stage two [1.4] (CEL06)

5.3.2 Prepositional phrase fragment formulaic sequences

Table 5.4 shows that Chinese EMI lecturers used exactly the same proportion of PP-based FS types as the native English lecturers, but significantly less FS tokens than the latter (G2=-88.52, p<0.0001). Table 5.6 lists the top 20 most frequent PP-based FSs in the two corpora.

Table 5.6 Top 20 most frequent PP-based FSs in CCL and CEL

Rank	CCL			CEL		
	FS	Freq.	Lectures	FS	Freq.	Lectures
1	*for example*	60	9	*of course*	67	11
2	*a lot of*	16	4	*a lot of*	58	11
3	*in that case*	16	2	*in terms of*	52	10
4	*this kind of*	13	5	*for example*	45	9
5	*according to*	11	5	*in fact*	40	11
6	*kind of*	10	2	*in other words*	33	7
7	*different kind of*	9	3	*kind of*	28	7
8	*because of*	9	3	*due to*	20	4
9	*such as*	6	4	*at least*	19	7
10	*a kind of*	6	4	*such as*	17	5
11	*due to*	5	4	*in order to*	17	5
12	*three kind of*	5	3	*lots of*	13	7
13	*of course*	4	4	*at the end of*	13	6

Table 5.6 Top 20 most frequent PP-based FSs in CCL and CEL (Cont.)

Rank	CCL			CEL		
	FS	Freq.	Lectures	FS	Freq.	Lectures
14	*another kind of*	4	2	*instead of*	12	8
15	*in the ocean*	3	2	*sort of*	12	6
16	*in the air*	3	2	*because of*	12	4
17	*in another way*	3	2	*in the case of*	12	2
18	*at the beginning*	3	2	*a couple of*	11	5
19	*in this picture*	2	2	*a number of*	11	4
20	*in this part*	2	2	*a little bit of*	11	3
	in this chapter	2	2			
	in this case	2	2			
	in terms of	2	2			
	different types of	2	2			
	at this time	2	2			
	at that time	2	2			
	all kind of	2	2			

In can be seen that the PP-based FSs are quite versatile, denoting time, range, place, degree, manner, amount, reason, etc. The most striking difference between CCL and CEL lecturers is related to FSs that encapsulate the type/kind phrases. The Chinese lecturers used seven such type/kind-related FSs, most of which refer to type/kind *per se*, e.g., *three kind of*, *another kind of*, *different types of*, *all kind of*. As discussed in Section 5.2, the omission of plural *-s* in *three kind of* and *all kind of* do not necessarily cause disturbance in communication. On top of that, the Chinese lecturers used the type/kind-related FSs *this kind of* and *a kind of* to provide examples, as exemplified in excerpt 111) and 112).

111) *But once you do **this kind of** fracture, the flow is become very, this is*

flow to, to your fracture, it's become linear flow, sometimes it was linear

flow inside your fracture. (CCL08)

112) *That's also **a kind of** igneous rocks, intrusive igneous rocks (CCL02)*

The Chinese lecturers also used the hedge metadiscourse *kind of* to withhold complete commitment to an assertion they make.

113) *So there are actually rocks, they have **kind of** similar compo- oh, sorry,*

compositions to the other extrusive igneous rocks (CCL02)

In except 113), the lecturer used *kind of* to imply to the students that what he was saying was his opinion instead of a solid fact.

On the other hand, the native English lecturers used only two such type/kind-related FSs, i.e., *kind of* and *sort of*. Functionally, they were also used as hedges to withhold complete commitment to assertions.

114) *er his experiments with rather more [0.7] bizarre [0.3] types of gauge*

*which might be [0.6] er **sort of** gauges of the future [10.6] (CEL04)*

In excerpt 114), the lecturer judged the gauge as 'gauges of the future', though in fact the gauge in the experiment may not be so bizarre as to be conceptualized something of the future.

In addition, the native English lecturers also used the vague expression *a little bit of* to downplay the force/importance of the argument 115), or simply to express a lesser level of degree 116).

115) *clearly you only have to stop and give **a little bit of** [0.5] social thought*

to this to realize these are numbers that have profound [0.4] social

consequences [1.0] (CEL06)

116) *so the mother [0.3] will potentially become sensitized [0.4] to the*
paternal antigens of the fetus now there is [0.3] ***a little bit of*** *exchange*
[0.3] of [0.2] blood [0.3] across the placenta [0.3] (CEL07)

Table 5.6 also shows that CCL and CEL lecturers used different range-denoting FSs. While the Chinese lecturers used six different types of FSs, the native English lecturers used two. Chinese lecturers used these FSs to either refer to a broad range (e.g., *in this picture*, *in this part*, and *in this chapter*) or specify a narrow scope (e.g., *in terms of*, *in that case*, *and in this case*), whereas the native English lecturers only used these FSs to refer to specified range (e.g., *in terms of* and *in the case of*).

117) *Now* ***in this chapter****, we'll study the derivative of differentiability.*
(CCL11)

118) *now the biggest error [0.6]* ***in terms of*** *measuring rainfall is [0.4] is is*
windspeed (CEL04)

Excerpt 117) sets 'derivative of differentiability' at the broad range of the chapter, while 118) confines the biggest error to those occurring when measuring rainfall. A query of the whole PP-based FS list reveals no broad range indicating FS use among native English lecturers.

Another difference lies in FSs that specify the amount. While the native English lecturers used *a lot of*, *lots of*, *a couple of*, and *a number of* to designate the

amount, the Chinese lecturers only used *a lot of* to denote large quantities, an indication of limited PP-based FS repertoire to denote the concept of amount.

5.3.3 Clausal phrase fragment formulaic sequences

Table 5.4 also shows that Chinese EMI lecturers used proportionally more Clausal FS types than the native English lecturers, but the actual frequency of use didn't show significant difference between them (G2=-0.24, p>0.05). Table 5.7 lists the top 20 most frequent clausal FSs in the two corpora.

Table 5.7 Top 20 most frequent Clausal FSs in CCL and CEL

Rank	CCL			CEL		
	FS	Freq.	Lectures	FS	Freq.	Lectures
1	*we can see*	79	8	*there are*	97	12
2	*you can see*	67	5	*there's*	92	11
3	*that means*	60	6	*I think*	62	11
4	*there is*	34	8	*I mean*	39	8
5	*Let's see*	33	3	*there is*	35	8
6	*there are*	27	6	*you know*	32	7
7	*Let's look at*	20	4	*you can see*	31	8
8	*you know*	18	6	*I'm gonna*	25	3
9	*That's why*	17	3	*what happens*	22	6
10	*I think*	14	5	*we're going to*	21	8
11	*do you know*	10	4	*we're gonna*	19	4
12	*We're going to talk about*	10	2	*as I said*	18	8
13	*we see*	7	3	*I'm not going to*	15	5
14	*we will talk about*	6	4	*I'm going to*	15	5
15	*I mean*	6	4	*that means*	14	6
16	*it means*	6	3	*there were*	13	4
17	*I show you*	6	2	*I hope*	12	6
18	*we will see*	5	3	*it doesn't matter*	12	3

Table 5.7 Top 20 most frequent Clausal FSs in CCL and CEL (Cont.)

Rank	CCL			CEL		
	FS	Freq.	Lectures	FS	Freq.	Lectures
19	*we are going to talk about*	5	2	*that's why*	11	4
20	*we are going to*	5	2	*as I say*	11	3
	if we want to	5	2			

Although the Chinese lecturers and the native English lecturers didn't show significant difference in terms of actual tokens of clausal FSs, the two groups of lecturers did have different preferences in the use of clausal FSs. The most striking difference lies in the use of FSs with pronouns 'we' and 'I', which are mainly in patterns of '**Let's + VP**', '**we + VP**', '**if + we + VP**', and '**(as) + I + VP**'. As has been examined in Section 5.2, Chinese lecturers tend to use the 'we' FSs to state aims and objectives, introduce topics, engage students, and explain concepts, while the native English lecturers preferred 'I' FSs to signpost the process of the lecture, give instructions, and make announcements. Nevertheless, Chinese lecturers also used 'I' FSs to either signpost the process of the lecture, or establish the lecturers' authority. Meanwhile, native English lecturers used negative 'I' FSs to help pin down and/or clarify the topic.

Apart from that, the lecturers also used many instances of FSs that incorporate '***there + be***' patterns. The CCL and CEL lecturers both used singular and plural forms of 'there + be' FSs. However, concordance line analysis reveals different usages between them in the 'there + are' FSs. In most cases, the Chinese lecturers would utter a specific numerical number after 'there + are', whereas the native English lecturers might have a wider varieties of words for them to use, e.g., *a number of*, *a*

whole of bunch of, *all kinds of*, *lots of*, and *a couple of*. This might be related to their different attitudes towards uncertainty of information. It is likely that Chinese lecturers might prefer to teach students only the knowledge in which they have complete confidence.

119) *Er,* ***there are*** <u>three</u> *kind of rock types in our nature. The igneous rocks, sedimentary rocks and metamorphic rocks*. (CCL05)

120) *um... in, this time period of, environmental awareness, and the more recent one of, of global biodiversity protection,* ***there are*** <u>a couple of</u> *other pieces, important pieces of legislation one is called CITES, that's the, acronym, and, that stands for Convention in International Trade in Endangered Species* (CEL09)

5.3.4 Verb phrase fragment formulaic sequences

Table 5.4 reveals that Chinese EMI lecturers used proportionally less VP-based FS types than the native English lecturers, and significantly less (G2=-159.7, p<0.0001) tokens than the latter. This is in sharp contrast with previous studies that investigated lexical bundles in academic writing (Bychkovska & Lee, 2017; Chen & Baker, 2010; Pan et al., 2016, Lu & Deng, 2019), which found much more use of VP-based bundles among non-native English speakers than the native English speakers. This opposite pattern in VP-based FS use could be due to the difference in the mode of the data, and most importantly a result of the limited FS repertoire of Chinese EMI lecturers. Table 5.8 lists the top 20 most frequent VP-based FSs in the two corpora.

Table 5.8 Top 20 most frequent VP-based FSs in CCL and CEL

Rank	CCL			CEL		
	FS	Freq.	Lectures	FS	Freq.	Lectures
1	look at	27	2	have to	89	11
2	have to	11	5	look at	55	10
3	depends on	7	3	go through	18	6
4	based on	6	2	is going to	17	6
5	figure out	5	2	tend to	15	7
6	belong to	4	3	looking at	15	6
7	is depends on	4	2	talk about	14	7
8	refer to	3	2	is going to be	13	5
9	fall into	3	2	work out	13	4
10	belongs to	3	2	depends on	12	6
11	tend to be	2	2	looked at	11	6
12				deal with	11	6
13				are going to	8	4
14				think about	8	3
15				worry about	7	5
16				cope with	7	4
17				set up	7	3
18				excuse me	7	3
19				comes from	7	2
20				depending on	6	6
				based on	6	6
				think of	6	3
				so to speak	6	3
				keep in mind	6	2

As shown in Table 5.8, most of the VP-based FSs are in the pattern of '***VP (-ed/ing) + PP***', e.g., *based on*, *cope with*, and *worry about*. Semantically, these VP-based FSs express diverse meaning categories, ranging from various colloquial forms denoting topic introduction/discussion, e.g., *look at* and *talk about*, to possession, e.g., *belong to*.

One thing worth noticing is the use of various colloquial forms of discussion. The Chinese lecturers used two forms of *look at* and *figure out* whereas the native English lecturers used seven different linguistic patterns to express the same concept, i.e., *deal with*, *look at, looked at, looking at*, *talk about*, *work out*, and *worry about*, which may again be evidence of limited FS repertoire among the Chinese EMI lecturers.

Another observation is the use of hedge metadiscourse. Both Chinese lecturers and native English lecturers used *tend to be* to express their personal evaluation to convey inexactitude 121-122).

121) *And the rocks made from andesite **tend to be** fine-grained.* (CCL02)

122) *the total clutch is uh, bigger than that. yeah and in in um, cases where it's been looked at, uh, looked at closely, there does **tend to be** an increase in uh, clutch size with female age.* (CEL11)

Apart from that, the native English lecturers also used *so to speak* to convey the inexactitude 123-124).

123) *um, so, the the communication network and the transportation netw-network, was really critical, um in terms of the demise, **so to speak** of the Passenger Pigeon.* (CEL09)

124) *or i believe i'm (uh) participating in the experiment, you may also find yourself in in collateral damage **so to speak**, so i'd like to introduce Bret and Janine, who just want to, fill you in and tell you about the form you've been handed out.* (CEL08)

The next thing to examine is the use of the idiomatic phrase *excuse me*. Though no such FS was found among Chinese lecturers, the native English lecturers used *excuse me* to either extend their apology for any annoyance caused 125) or more importantly as an alternative way to self-correct 126).

125) *mkay so this is an example how this kind of analysis can help you <COUGH>* **excuse me** *help you evaluate different different management options*. (CEL11)

126) *so that's a three percent increase in survival, uh* **excuse me** *three percent increase in population growth rate from that plan*. (CEL11)

Table 5.8 also reveals the exclusive use of the FS *keep in mind* among the native English lecturers. *Keep in mind* is often used to provide caveats 127) or remind students of students of class-related issues 128), which might help improve classroom teaching effectiveness.

127) *and* **keep in mind** *you're not just trying to break the rack up into regions anymore you're looking at two points at a time*. (CEL10)

128) *okay now, what i want you to* **keep in mind** *is 'm gonna go through i talked about those different stages* (CEL09)

5.3.5 Conjunction phrase fragment formulaic sequences

Table 5.4 indicates that both Chinese EMI lecturers and the native English lecturers used very small proportions of ConjP-based FS types, meanwhile there is no significant difference between them in terms of the actual occurrence of this category

of FSs (G2=+0.16, p>0.05). Nevertheless, they still worth careful examination since conjunction phrase fragment formulaic language often involve discontinuous FSs, which were seldom investigated in precious studies. Table 5.9 lists the top 20 most frequent ConjP-based FSs in CCL and CEL.

Table 5.9 Top 20 most frequent ConjP-based FSs in CCL and CEL

Rank	CCL			CEL		
	FS	Freq.	Lectures	FS	Freq.	Lectures
1	*so that*	26	2	*so that*	69	11
2	*because ⋯ so*	13	6	*as soon as*	4	3
3	*not only ⋯ but also*	3	3	*as long as*	3	3
4	*if ⋯ then*	3	2	*what if*	3	2
5				*even though*	3	2
6				*not only ⋯ but*	2	2
7				*not ⋯ but*	2	2

Only four and seven ConjP-based FSs were found in CCL and CEL, respectively. As shown in Table 5.9, *so that* tops both CCL and CEL. It was generally used by both Chinese and native English lecturers to express intentions, as illustrated in 129-130).

129) *So one reason is we want to extend a conductive path deep into a formation, **so that** we can increase productivity beyond the natural level*. (CCL08)

130) *so of course one of your strategies [0.3] may be [0.2] to put some adaptability into your breeding programme **so that** you can [0.2] move*

in different directions not just [0.3] selecting for a single [0.5] er [0.2]

particular picture [1.9] (CEL05)

In addition, Chinese lecturers used the non-standard *because ... so* to express cause-effect relations, which could be attributed to the logical relation expressions in Chinese, where both the concepts of cause (Chinese: yinwei) and effect (Chinese: suoyi) are verbally expressed concurrently. As discussed earlier, such non-standard forms do not cause any disturbance or miscommunication.

Table 5.9 also shows that Chinese lecturers used *not only ... but also* 131) to express coordinating relations while the native English lecturers tended to use the shortened form *not only ... but* 132).

131) *Its purpose is **not only** to restore formation pressure, **but also** to improve oil displacement or fluid flow in the reservoir.* (CCL10)

132) *the period of time pre-eighteen-fifty, is thought of **not only** for birds **but** for all, uh, essentially uh, wildlife species, as a period of abundance.* (CEL09)

Besides, some of the ConjP-based FSs were used to express conditional relations. Chinese lecturers used *if ⋯ then* 133) to introduce potential consequences on the basis of assumptions while native English lecturers preferred to use *as soon as* 134) and *as long as* 135).

133) ***if** it is come from water, **then** the engineer has to figure out something to do it.* (CCL06)

134) *at that instant [0.3] course **as soon as** it rolls a little distance [0.8] the point of instantaneous centre is going to move round the rim* (CEL02)

135) *this bird is found in Asia it's not a native species, and, um it does well, um in many agricultural, regions **as long as** there's, good cover and not, uh s- harsh winters*. (CEL09)

The next observation is that native English lecturers used *what if* to engage students in further thinking through extreme conditions 136-137).

136) *but **what if** we could actually shorten the pulse further [0.6] would we see earlier events after all in the track model [0.2]* (CEL01)

137) *but again there's a massive ethical problem here which doesn't seem to have been recognized [0.5] is that **what if** [0.5] subsequently that patient [0.5] er [0.5] acute rejection or chronic rejection occurred [0.5] and the face [0.2] [0.6] was rejected what do you do you nip out and find another corpse [1.2]* (CEL07)

Finally, Table 5.9 also shows the exclusive use of FSs by the native English lecturers, with *even though* denoting concession 138), and *not ... but* clarifying topics and/or stressing their points 139).

138) *here in the Great Lakes, **even though** there are fewer plovers, this bird is a high priority because of its uniqueness* (CEL09)

139) *and um, CITES is, one of the strongest um, as as i said **not** a piece of legislation **but** uh strongest um agreement that helps to, to curb um,*

trade in, in parrots. (CEL09)

5.4 An ELF perspective of formulaic sequences in CCL and CEL

Since formulaic sequences allow both linguistic convention and creativity (Mauranen, 2009), the present study also investigates FSs from the ELF perspective. A total of 34 non-standard FS types and 63 non-standard FS tokens were identified in CCL lectures, accounting for 6% of the total FS types and 4% of the total FS tokens respectively. Table 5.10 presents the overall statistics of non-standard FSs in CCL.

Table 5.10 Overall statistics of non-standard FSs in CCL

Lecture	Non-standard FS types	Non-standard FS tokens
CCL01	9	11
CCL02	2	9
CCL03	5	8
CCL04	2	2
CCL05	4	6
CCL06	4	5
CCL07	3	3
CCL08	5	7
CCL09	1	1
CCL10	1	1
CCL11	2	9
CCL12	1	1
Total	34	63

With reference to Seidlhofer (2004), the present study summarized nine different categories of non-standard formulaic sequences in terms of the lexico-grammatical features, which are presented as in Table 5.11.

Table 5.11 Non-standard FS categories in CCL

Non-standard categories	FS	Freq.	Texts
1) Inserting redundant copula *be*	*It's mean*	8	1
	is depends on	4	2
	It's looks like	2	1
	is belong to	2	1
	was depends on	1	1
	It's means	1	1
	It's depend on	1	1
	it is means	1	1
	is strongly depends on	1	1
	is depend on	1	1
	is belongs to	1	1
2) Inserting redundant conjunction	*because ... so*	12	5
3) Inserting redundant prepositions	*We'll discuss about*	1	1
4) Misusing or omitting prepositions	*another way*	3	1
	for this pictures	2	1
	caused from	2	1
	the storage for	1	1
	origin from	1	1
	have effect to	1	1
	distinguish it with	1	1
	concentrate in	1	1
	by the name of	1	1
	another side	1	1
	a waste for	1	1
5) Omitting obligatory or inserting redundant plural *-s*	*Let's see this pictures*	3	1
	two kind of	1	1
5) Omitting obligatory or inserting redundant plural *-s*	*lot of*	1	1
	Let's look at this pictures	1	1
	all kind of	1	1
6) Omitting obligatory or inserting definite and indefinite articles	*in the recent years*	1	1
7) Misusing 3rd person singular *-s*	*this picture show*	1	1
8) Failing to use nominalized verb forms	*with basic assume*	1	1
	the begin of	1	1
9) Failing to use existential constructions	*there has*	1	1

Björkman (2008b) classified the non-standard FSs into three subcategories, including: 1) approximated FSs that lead to disturbance; 2) approximated FSs that lead to no disturbance; and 3) approximated FSs that increase comprehensibility. Concordance analysis indicates that none of the identified non-standard FSs in CCL leads to communication breakdowns.

In addition, most of the non-standard FSs were actually quite idiosyncratic since they were mostly used by only one lecturer in the corpus. Three common non-standard FSs used by CCL lecturers include *because … so,* approximated forms of *depend on* (*is depends on*, *was depends on*, *is strongly depends on*, *is depend on*, and *it's depend on*), and approximated forms of *belong to* (*is belong to* and *is belongs to*). These non-standard FSs are semantically straightforward, which do not seem to cause any confusion or disturbance in communication.

5.5 Summary

This chapter reports the results and discussion of formulaic sequences in EMI lectures in CCL and CEL.

A total of 1615 and 4354 formulaic sequences were identified in CCL and CEL lectures, respectively. The log likely-hood comparisons revealed significantly less use of FSs in terms of both types and tokens by the Chinese EMI lecturers than their native English counterparts. It was speculated that Chinese EMI lecturers might possess a more limited repertoire of FSs than the native English lecturers, which was corroborated by the interview data.

In order to give a quick glance at the formulaic sequences used in the two corpora, both frequency and dispersion were adopted to extract the most frequent formulaic sequences. Ten of the top 20 most frequent FSs were shared by the two groups of lecturers, which are either verb-related phrases or clausal phrase fragments. A notable difference between CCL and CEL lecturers lies in the use of 'we' FSs and 'I' FSs. Chinese EMI lecturers used significantly more 'we' FSs to state aims and objectives, introduce topics, engage students, and explain concepts whereas the native English lecturers used significantly more 'I' FSs to signpost the process of the lecture or establish the lecturers' authority.

On the basis of Biber et al (1999) and Wang (2017), the present study classified the FSs into eight different structural categories, including NP fragment, PP fragment, AP fragment as modifier, AP fragment as complement, Clausal fragment, VP fragment, AdP fragment, and ConjP fragment. Both CCL and CEL lectures exhibit a mixed feature of 'oral' and 'literate' characteristics. In term of token frequencies, lecturers mainly relied on NP-, PP-, clausal, VP-, and ConjP-based FSs.

Structurally, most of the NP-based NPs are in the pattern of '***the + noun + of***'. Chinese EMI lecturers used significantly less NP-based FSs than native English lecturers. Meanwhile, their use of such FSs are less diverse FSs than their English counterparts. Besides, vague expressions such as *things like that* and *something like this*, and FSs that incorporate shell nouns are generally lacking among Chinese EMI lecturers.

The most striking difference in the use of PP-based FSs between CCL and CEL lecturers is related to FSs that encapsulate the type/kind phrases. Chinese lecturers used seven such type/kind-related FSs, most of which refer to type/kind *per se*, e.g., *three kind of*, *another kind of*, *different types of*, *all kind of*. Other type/kind-related FSs (*this kind of* and *a kind of*) were used by Chinese lecturers to provide examples. The last type/kind-related FS *kind of* was used by Chinese lecturers as hedges to withhold complete commitment to assertions. On the other hand, native English lecturers only used two type/kind-related FSs, i.e., *kind of* and *sort of*, both serving as hedges.

On top of that, lecturers in CCL and CEL also differ in PP-based FSs denoting range. While the Chinese lecturers used six different types of FSs, the native English lecturers used two. Chinese lecturers used these FSs to either refer to a broad range (e.g., *in this picture*, i*n this part*, and *in this chapter*) or specify a narrow scope (e.g., *in terms of*, *in that case*, and *in this case*), whereas the native English lecturers only used these FSs to refer to specified range (e.g., *in terms of* and *in the case of*).

Most Clausal FSs include pronouns as clause subjects, and are mainly in the patterns of '**Let's + VP**', '**we + VP**', '**if + we + VP**', and '**(as) + I + VP**'. Though Chinese EMI lecturers and native English lecturers didn't show significant difference in the number of Clausal FSs, they did exhibit different preferences in the choice of pronoun-embedding Clausal FSs.

Most of the VP-based FSs are in the pattern of 'VP (-ed/ing) + PP', e.g., *based on*, *cope with*, and *worry about*. One striking difference between Chinese and English EMI

lecturers lies in the use of FSs denoting the concept of discussion. Chinese lecturers used two forms of FSs (i.e., *look at* and *figure out*) to express the notion of discussion whereas the native English lecturers used seven different forms (i.e., *deal with*, *look at*, *looked at*, *looking at*, *talk about*, *work out*, *worry about*) to express the same concept, which may again be evidence of limited FS repertoire of the Chinese EMI lecturers.

Only four and seven ConjP-based FSs were found in CCL and CEL, respectively. Topping the ConjP-based FS list of CCL and CEL lectures, *so that* was generally used to express intentions. Besides, Chinese lecturers used the non-standard *because ... so* to express cause-effect relations, which could be attributed to the logical relation expressions in Chinese, where both the concepts of cause (Chinese: yinwei) and effect (Chinese: suoyi) are verbally expressed concurrently.

An ELF perspective of analysis identified 34 non-standard FS types and 63 non-standard FS tokens by the CCL lecturers, accounting for 6% of the total FS types and 4% of the total FS tokens, respectively. These non-standard FSs were summarized into nine different categories in terms of their lexico-grammatical features. In congruence with Björkman (2008b), none of them seem to cause any disturbance in communication. In addition, most of the non-standard FSs were actually quite idiosyncratic except for three common ones, viz., *because ... so,* approximated forms of *depend on* (*is depends on*, *was depends on*, *is strongly depends on*, *is depend on*, and *it's depend on*), and approximated forms of *belong to* (*is belong to* and *is belongs to*).

CHAPTER 6

RESULTS AND DISCUSSION III: CONNECTING

FORMULAIC SEQUENCES AND

RHETORICAL MOVES

This chapter examines the functions of formulaic sequences in connection with rhetorical moves/steps where they are situated. It first presents the overall statistics of formulaic language functions in CCL and CEL lectures. Then it elaborates move signaling formulaic sequences in individual moves/steps in different lecture phases. Finally, it ends with a summary of the chapter.

6.1 Overview of formulaic language functions in CCL and CEL

The results and discussion on formulaic language functions in CCL and CEL lectures are expected to answer research question: 3) What are the potential connections between formulaic sequences and rhetorical move structure?

Table 6.1 Statistics of FS functions in CCL and CEL

FS functions	CCL (%)	CEL (%)	LL critical value G2[a]
MS	338 (21%)	794(18%)	-13.47***
MB	1032 (64%)	3345 (77%)	-267.46****
MN	245 (15%)	215 (5%)	+63.94****
Total	1615 (100%)	4354 (100%)	-172.37****

Note. [a] The log-likelihood (LL) critical value G2 itself is always a positive number. However, indicators

'+' and '-' are inserted to show overuse and underuse of corpus 1 (in this case CCL) relative to corpus 2 (in this case CEL). *p < 0.05, critical value G2= 3.84; **p < 0.01, critical value G2 = 6.63; ***p < 0.001, critical value G2 = 10.83; ****p < 0.0001, critical value G2 = 15.13.

Table 6.1 presents statistics of move-formulaicity connections, i.e., FS functions in rhetorical moves/steps, in CCL and CEL. The table shows that, in terms of proportion, Chinese EMI lecturers used slightly more **Move Signaling formulaic sequences** (MS FSs) and more **Move Neutral formulaic sequences** (MN FSs), but less **Move Building formulaic sequences** (MB FSs) than the native English lecturers. It can be seen the majority of FSs in the two corpora (CCL: 85% and CEL: 95%) are directly or indirectly connected with rhetorical moves.

In terms of actual occurrences, Chinese lecturers used significantly less MS FSs (G2=-13.47, p<0.001) and MB FSs (G2=-267.46, p<0.0001) than the native English lecturers, but significantly more MN FSs (G2=+63.94, p<0.0001) than native English lecturers. It seems that Chinese lecturers might be less equipped with move-related FSs, i.e., MS FSs and MB FSs, therefore they resorted to MN FSs. In order to be of direct pedagogical value, the present study focuses on MS FSs only. The following parts investigate MS FSs in each move/step in EMI lectures in CCL and CEL.

6.2 Move signaling formulaic sequences in moves/steps of the Opening Phase

Table 6.2 shows the statistics of MS FSs in the moves/steps of the Opening Phase.

Table 6.2 Statistics of MS FSs in the Opening Phase

Move	MS FSs		
	CCL	CEL	LL critical value G2[a]
M1	1	3	-0.19
M2S1	0	0	0
M2S2	1	6	-1.53
M2S3	0	2	-1.72
M2S4	1	13	-6.09*
M3S1	12	16	+0.74
M3S2	26	10	+20.67****

Note. [a] The log-likelihood (LL) critical value G2 itself is always a positive number. However, indicators '+' and '-' are inserted to show overuse and underuse of corpus 1 (in this case CCL) relative to corpus 2 (in this case CEL). *$p < 0.05$, critical value G2= 3.84; **$p < 0.01$, critical value G2 = 6.63; ***$p < 0.001$, critical value G2 = 10.83; ****$p < 0.0001$, critical value G2 = 15.13.

As revealed in table 6.2, there is no significant difference in the use of MS FSs in most moves/steps in the Opening Phase between the two corpora. However, Chinese lecturers used significantly less MS FSs in *M2S4* (G2=-6.09, p<0.05), but significantly more MS FSs in *M3S2* (G2=+20.67, p<0.0001) than the native English lecturers.

Table 6.3 MS FSs in *Move 1*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M1	1	*Let's get it started*	1	1	*shall we start*	1
				2	*Let's start*	1
				3	*why don't we get started*	1

Table 6.3 shows the MS SFs (e.g., *Let's get it started* and *shall we start*) that CCL

and CEL lecturers use in **M1 Getting Started** (See excerpt 140-141)).

140) <M1>S1: OK, ***let's get it started***. Good afternoon, everyone!</M1> (CCL08)

141) <M1>okay ***let's start*** uh, i think it's just right on time.</M1> (CEL10)

According to table 6.2 and table 6.3, there is no significant difference in the use of

MS FSs in **M1 Getting Started** between Chinese and native English lecturers, but the

latter seem to have richer inventories of MS FSs to signpost the beginning of the lecture.

Table 6.4 MS FSs in *Move 2*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M2S1			0			0
M2S2	1	*we talked about*	1	1	*what I said*	1
				2	*we've been talking about*	1
				3	*we talked about*	1
				4	*had looked at*	1
				5	*bring you back to*	1
				6	*at the end of*	1
M2S3			0	1	*you're gonna see*	1
				2	*I'm going to*	1
M2S4	1	*If … please*	1	1	*you'll see*	2
				2	*what I wanted to say*	1
				3	*we'll come back to*	1
				4	*I'm not going to*	1
M2S4				5	*I'm just gonna*	1
				6	*I'm gonna*	1
				7	*I'm going to*	1
				8	*I'd like to*	1
				9	*feel free to ask*	1
				10	*are there any questions*	1
				11	*any questions*	1
				12	*any comments*	1

Table 6.4 presents the MS SFs (e.g., *we talked about*, *what I said* and *you're gonna*

see) that CCL and CEL lecturers use in different steps of ***M2 Warming Up*** (See excerpt

142-143)).

142) <M2S2>Let's see, last class in this section chapter two, ***we talked about***

conventional artificial lift methods, such as like sucker rod pumps, ESP, Gas

lift and PCP.</M2S2> (CCL06)

143) <M2S4>okay [0.4] now if you do have questions then please do ***feel free to***

ask them as we [0.2] go along and i will try and remember at the end of each

lecture if i don't shout at me [0.3] to please ask whatever questions you want

to okay [0.5] but [0.4] otherwise just ask as we …</M2S4> (CEL05)

According to table 6.2 and table 6.4, Chinese EMI lecturers seem to have used less

MS FSs to signify individual steps of ***M2 Warming Up*** than the native English lecturers.

Though significantly more step ***M2S1*** has been identified in EMI lectures in CCL than

in CEL (refer to Section 4.2), no MS FS has been found in ***M2S1***. This could be due to

Chinese EMI lecturers' incapability of using MS FSs to signal the rhetorical functions

of ***M2S1***.

The two tables also show more use of MS FS in other steps of ***M2*** by lecturers in

CEL, ***M2S4*** in particular (G2=-6.09, p<0.05). That might be attributed to the native

English lecturers' inclination to use step ***M2S4*** to help inform students of lecturers'

intentions and maintain close teacher-student relationship, as suggested in Section 4.2.

Table 6.5 MS FSs in *Move 3*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M3S1	1	*we'll study*	2	1	*we're going to*	3
	2	*we are going to*	2	2	*talk about*	2
	3	*I will talk about*	2	3	*what I'm going to*	1
	4	*we're going to talk about*	1	4	*what I want to talk about*	1
	5	*we'll learn*	1	5	*what I want to do*	1
	6	*we will talk about*	1	6	*we're gonna*	1
	7	*we will learn*	1	7	*we [1.3] move on*	1
	8	*I will talk something about*	1	8	*start off by*	1
	9	*I will learn*	1	9	*let's actually get on with*	1
				10	*I'm gonna*	1
				11	*I'm going to*	1
				12	*I'd like to*	1
				13	*be talking about*	1
M3S2	1	*we'll study*	4	1	*go through*	3
	2	*and then we move to*	4	2	*what I'd like to do*	1
	3	*we'll talk about*	2	3	*what I want to cover*	1
	4	*we are going to talk about*	2	4	*we're gonna say*	1
	5	*we are going to*	2	5	*we're gonna assume*	1
	6	*we're going to talk about*	1	6	*I'm going to talk*	1
	7	*we'll go back*	1	7	*go on to*	1
	8	*we will roughly talk about*	1	8	*at the end*	1
	9	*we will learn*	1			
	10	*we talk about*	1			
	11	*then we move to*	1			
	12	*let's look at*	1			
	13	*in this part*	1			
	14	*in this chapter*	1			
	15	*I will show you*	1			
	16	*at the beginning*	1			
	17	*at last*	1			

Table 6.5 presents the MS SFs (e.g., *we'll study*, *we're going to* and *what I want to talk about*) that CCL and CEL lecturers use in different steps of ***M3 Setting Up Lecture Agenda*** (See excerpt 144-145)).

144) <M3S1>Then today ***we are going to*** start a new chapter. This chapter three, basically Advanced Hydraulic Fracturing Technology.</M3S1> (CCL08)

145) <M3S2>okay. well ***what i'd like to do*** first... and then when we look at the code, i think it'll be easier to understand if we've just looked at these, uh graphics again to see how the sort works.</M3S2> (CEL12)

Table 6.2 and table 6.5 reveal that Chinese EMI lecturers used significantly more MS FS tokens (G2=+20.67, p<0.0001) than the native English lecturers in ***M3S2***, which could be due to Chinese lecturers' inclination to include explicit proclamation of lecture scope for students, as suggested in the genre analysis of step ***M3S2*** in Section 4.2.

6.3 Move signaling formulaic sequences in moves/steps of the Theme Network Building Phase

Table 6.6 presents the statistics of MS FSs in the moves/steps of Theme Network Building Phase.

Table 6.6 Statistics of MS FSs in the Theme Network Building Phase

Move	MS FSs		
	CCL	CEL	LL critical value G2[a]
M4	50	86	+0.19
M5S1	1	4	-0.54
M5S2	7	6	+1.92
M5S3a	20	10	+12.42[***]

Table 6.6 Statistics of MS FSs in the Theme Network Building Phase (Cont.)

Move	MS FSs		
	CCL	CEL	LL critical value G2ᵃ
M5S3b	3	11	-1.23
M5S3c	1	4	-0.54
M5S4	40	74	+0.00
M5S5	87	280	-21.91****
M5S6	4	23	-5.56*
M5S7	9	63	-18.91****
M5S8	1	4	-0.54
M5S9	25	29	+2.92
M5S10	3	49	-25.57****
M6S1	3	21	-6.30*
M6S2	11	17	+0.22
M6S3	15	21	+0.69
M7	0	4	-3.45
M8	0	6	-5.17*
M9	5	14	- 0.66

Note. ᵃ The log-likelihood (LL) critical value G2 itself is always a positive number. However, indicators '+' and '-' are inserted to show overuse and underuse of corpus 1 (in this case CCL) relative to corpus 2 (in this case CEL). *p < 0.05, critical value G2= 3.84; **p < 0.01, critical value G2 = 6.63; ***p < 0.001, critical value G2 = 10.83; ****p < 0.0001, critical value G2 = 15.13.

As revealed in table 6.6, Chinese EMI lecturers used significantly less MS FSs in most moves/steps in this phase. Specifically, Chinese lecturers used significantly less MS FSs in *M5S5* (G2=-21.91, p<0.0001), *M5S6* (G2=-5.56, p<0.05,), *M5S7*(G2=-18.91, p<0.0001), *M5S10* (G2=-25.57, p<0.0001), *M6S1* (G2=-6.30, p<0.05), and *M8* (G2=-5.17, p<0.05), but significantly more MS FSs in *M5S3a* (G=+12.42, p<0.001).

Table 6.7 MS FSs in *Move 4*

Move		CCL			CEL	
	N	MS FS	tokens	N	MS FS	tokens
	1	*let's look at*	8	1	*we're going to*	7
	2	*let's see*	7	2	*look at*	5
	3	*we can see*	3	3	*talk about*	4
	4	*let's talk about*	3	4	*what about*	3
	5	*let's study about*	3	5	*think about*	3
	6	*we're going to talk about*	2	6	*let's look at*	3
	7	*we will look at*	2	7	*I'm gonna*	3
	8	*turn to*	2	8	*worry about*	2
	9	*what we should know*	1	9	*we move on to*	2
	10	*what I'm going to talk about*	1	10	*touch on*	2
	11	*we're going to*	1	11	*the question is*	2
	12	*we'll talk about it*	1	12	*let's go on*	2
	13	*we'll look at*	1	13	*I wanna*	2
	14	*we will talk about*	1	14	*what we're going to*	1
	15	*we will see*	1	15	*what we're doing*	1
M4	16	*we will learn*	1	16	*what we'd like to do*	1
	17	*we talk about*	1	17	*what we wanna do*	1
	18	*we are talk about*	1	18	*what I'm going to do*	1
	19	*we are going to talk about*	1	19	*what I'd like to do*	1
	20	*the next step*	1	20	*we're gonna see*	1
	21	*let's start*	1	21	*we're gonna*	1
	22	*let's learn*	1	22	*we're going to look at*	1
	23	*let's go on*	1	23	*we'll get on*	1
	24	*lets' er learn*	1	24	*we go on to*	1
	25	*let's continue*	1	25	*there's another thing*	1
	26	*let me just show you*	1	26	*there's another issue*	1
	27	*let me introduce*	1	27	*the question was*	1
	28	*do you know*	1	28	*the next thing*	1
				29	*take a look [1.0] at*	1
				30	*one of the questions*	1
				31	*one of the problems*	1
				32	*move along to*	1

Table 6.7 MS FSs in *Move 4* (Cont.)

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M4				33	*make the next argument*	1
				34	*let's talk about*	1
				35	*let's take a look at*	1
				36	*let's start looking at*	1
				37	*let's start [0.2] with*	1
				38	*let's look*	1
				39	*let's have a look*	1
				40	*let's go to*	1
				41	*let's go on to*	1
				42	*let's [0.2] move on*	1
				43	*let's [0.2] go on*	1
				44	*let's [0.2] [0.2] briefly look at*	1
				45	*let us think*	1
				46	*lemme point out*	1
				47	*lemme get back to*	1
				48	*I'm, gonna*	1
				49	*I'm just gonna*	1
				50	*I'm going to*	1
				51	*I'll talk about*	1
				52	*I'll show you*	1
				53	*if we're going to*	1
				54	*I want to talk about*	1
				55	*going back to*	1
				56	*go through*	1
				57	*go about [0.3] trying*	1
				58	*deal with*	1
				59	*come across*	1

Table 6.7 presents the MS SFs (e.g., *let's look at*, *let's see* and *we're going to*) that

CCL and CEL lecturers use in *M4 Introducing The Topic* (See excerpt 146-147) for

examples of MS FSs in *M4*).

146) <M4>So ***let's look at*** this different types of weathering. The first is physical

or mechanical weathering.</M4> (CCL05)

147) <M4>now, the next thing ***we're going to look at*** is the actual C-plus-plus code

for it,</M4> (CEL12)

Table 6.6 and table 6.7 show that there is no significant difference in the use of MS

FSs in ***Move 4 Introducing the topic***. However, the Chinese EMI lecturers used only

about half as many different types of MS FSs as the native English lecturers. Meanwhile,

since Chinese EMI lecturers used significantly more ***M4*** than their English counterparts

as revealed in Section 4.3, it might be speculated the former may have relied heavily

on limited MS FSs in ***M4***, an indication of Chinse lecturers' limited FS repertoire.

Table 6.8 MS FSs in *Move 5*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S1	1	*be defined as*	1	1	*we're gonna assume*	1
				2	*let me introduce*	1
				3	*it means*	1
				4	*by definition*	1
M5S2	1	*has to*	2	1	*are completely dependent on*	3
	2	*we'll talk about*	1	2	*'re not completely dependent on*	1
	3	*we are going to talk about*	1	3	*it's important*	1
	4	*that's why*	1	4	*have to*	1
	5	*I'm going to talk about*	1			
	6	*I'll talk about*	1			
M5S3a	1	*let's see*	5	1	*if you look at*	2
	2	*let's look at*	5	2	*the case of*	1
	3	*I show you*	2	3	*'re well familiar with*	1
	4	*we will talk about*	1	4	*it has been said*	1
	5	*we talk about*	1	5	*in many cases*	1

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	\multicolumn{3}{CCL}			\multicolumn{3}{CEL}		
	N	MS FS	tokens	N	MS FS	tokens
M5S3a	6	*we are going to talk about*	1	6	*if you're looking at*	1
	7	*refresh your memory*	1	7	*if you think about it*	1
	8	*I will tell you*	1	8	*if we look at*	1
	9	*do you think*	1	9	*heard of*	1
	10	*do you know*	1			
	11	*as I have said*	1			
M5S3b	1	*let's see*	1	1	*in the case of*	3
	2	*I'd like to*	1	2	*you've probably seen*	1
	3	*have you been to*	1	3	*what about*	1
				4	*under those circumstances*	1
				5	*let's think of*	1
				6	*in this case*	1
				7	*in the morning*	1
				8	*if you look at*	1
				9	*be familiar [0.6] with*	1
M5S3c	1	*at this point*	1	1	*with that in mind*	1
				2	*when we talk about*	1
				3	*it is the case [1.5] that*	1
				4	*if you remember*	1
M5S4	1	*look at*	8	1	*in terms of*	8
	2	*we're going to talk about*	3	2	*in other words*	6
	3	*the definition of*	3	3	*for example*	5
	4	*it's mean*	3	4	*in fact*	4
	5	*I show you*	3	5	*I think*	4
	6	*I'm going to talk about*	2	6	*for instance*	4
	7	*if we want to*	2	7	*the idea of*	3
	8	*what does it mean*	1	8	*such as*	2
	9	*we're going to*	1	9	*in the case of*	2
	10	*we will see*	1	10	*in practice*	2
	11	*we should know*	1	11	*I mean*	2
	12	*we see*	1	12	*what we're going to see*	1
	13	*that means*	1	13	*what we are assuming*	1
	14	*let's show you*	1	14	*what I'm trying to*	1

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S4	15	*let's see*	1	15	*what I mean*	1
	16	*let's look at*	1	16	*we're gonna assume*	1
	17	*let us see*	1	17	*the question is*	1
	18	*let us know*	1	18	*the problem is*	1
	19	*I will show you*	1	19	*the point that*	1
	20	*I can show you*	1	20	*the point is*	1
	21	*do you know*	1	21	*the idea was*	1
	22	*based on*	1	22	*the fact that*	1
	23	*according to*	1	23	*the assumption is*	1
				24	*that means*	1
				25	*one of the things*	1
				26	*make the assumption*	1
				27	*look at*	1
				28	*let's just think about*	1
				29	*let's assume*	1
				30	*it means that*	1
				31	*is defined as*	1
				32	*in the context*	1
				33	*in that situation*	1
				34	*in some cases*	1
				35	*in principle*	1
				36	*if you're looking at*	1
				37	*if we look at*	1
				38	*I should say*	1
				39	*for that matter*	1
				40	*for example*	1
				41	*focus on*	1
				42	*as a principle*	1
				43	*are [0.8] referred to*	1

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S5	1	*for example*	26	1	*for example*	18
	2	*that means*	12	2	*and so on*	15
	3	*it means that*	5	3	*in fact*	13
	4	*that's why*	4	4	*in other words*	12
	5	*such as*	4	5	*I mean*	10
	6	*let's see*	4	6	*I think*	9
	7	*let's look at*	3	7	*in terms of*	8
	8	*it means*	3	8	*you can see*	6
	9	*we can see*	2	9	*we're gonna*	6
	10	*we're going to*	1	10	*look at*	5
	11	*we'll look at*	1	11	*let's look at*	5
	12	*we will see*	1	12	*I'm gonna*	5
	13	*we study about*	1	13	*what happens*	4
	14	*we see here*	1	14	*think of*	4
	15	*we can look at this*	1	15	*that's why*	4
	16	*to be more specific*	1	16	*in the case of*	4
	17	*this table shows*	1	17	*as you can see*	4
	18	*this picture shows*	1	18	*what you can see*	3
	19	*this picture show*	1	19	*we'll look at*	3
	20	*the problem is*	1	20	*that means*	3
	21	*look at*	1	21	*let's say*	3
	22	*let's take the rock cycle for instance*	1	22	*it's estimated that*	3
	23	*let's see this the example*	1	23	*you'll see*	2
	24	*let's make an example*	1	24	*with respect to*	2
	25	*let's look at this pictures*	1	25	*what we're doing*	2
	26	*let's go back to*	1	26	*we're looking at*	2
	27	*let me give you*	1	27	*we can see*	2
	28	*it is means*	1	28	*we are going to*	2
	29	*in fact*	1	29	*think about*	2
	30	*I'm going to*	1	30	*the definition of*	2
	31	*if we want to*	1	31	*refer to this as*	2
	32	*I mean*	1	32	*looking at*	2

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S5	33	*because of*	1	33	*let's go back to*	2
				34	*let us suppose*	2
				35	*I'm gonna show you*	2
				36	*I'm going to*	2
				37	*if you imagine*	2
				38	*I should say*	2
				39	*I believe*	2
				40	*for instance*	2
				41	*concentrate on*	2
				42	*you'll notice*	1
				43	*you would say*	1
				44	*you know*	1
				45	*when we talked about*	1
				46	*what's happened*	1
				47	*what we're saying*	1
				48	*what we're really doing*	1
				49	*what we have to do*	1
				50	*what we expect to see*	1
				51	*what we can see*	1
				52	*what we are going to be doing*	1
				53	*what this meant*	1
				54	*what this means*	1
				55	*what it means*	1
				56	*what I'm gonna*	1
				57	*what I'm doing*	1
				58	*what I mean*	1
				59	*what am I talking about*	1
				60	*we're thinking of*	1
				61	*we're gonna use*	1
				62	*we're about to*	1
				63	*we expect to see*	1
				64	*we can see that*	1
				65	*we can also see*	1

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S5				66	*this issue of*	1
				67	*the trouble is*	1
				68	*the reason for that*	1
				69	*the question is*	1
				70	*the problem of*	1
				71	*the problem is*	1
				72	*the key feature to*	1
				73	*the issue of*	1
				74	*the idea of*	1
				75	*the first thing to*	1
				76	*the first point to note*	1
				77	*that is why*	1
				78	*such as*	1
				79	*start off with*	1
				80	*referring to*	1
				81	*refer to as*	1
				82	*refer to*	1
				83	*one thing to mention*	1
				84	*one of the things*	1
				85	*let's think about*	1
				86	*let's then see*	1
				87	*let's take a look at*	1
				88	*let's take a look again at*	1
				89	*let's start off with*	1
				90	*let's see*	1
				91	*let's just take a quick look at*	1
				92	*let's just say*	1
				93	*let's have a quick look*	1
				94	*let's have a look [0.9] at*	1
				95	*let's go back*	1
				96	*let's also suppose*	1
				97	*let's [0.4] look at*	1
				98	*let me see*	1

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
				99	*let me get to that*	1
				100	*lemme go back to*	1
				101	*it's estimated*	1
				102	*it turns out that*	1
				103	*it stacks up to*	1
				104	*it means*	1
				105	*is looking at*	1
				106	*is concerned about*	1
				107	*in virtually all case*	1
				108	*in this case*	1
				109	*in theory*	1
				110	*in that case*	1
				111	*in terms [0.3] of*	1
				112	*in some cases*	1
				113	*in principle*	1
				114	*in particular*	1
M5S5				115	*in many cases*	1
				116	*in effect*	1
				117	*in all cases*	1
				118	*I'm looking for*	1
				119	*if you think of*	1
				120	*if you think about it*	1
				121	*if you look at*	1
				122	*if you look*	1
				123	*if we try to*	1
				124	*if we look at*	1
				125	*if we imagine*	1
				126	*if we assume*	1
				127	*I would like you to*	1
				128	*I suppose*	1
				129	*here you see*	1
				130	*focused on*	1
				131	*does anyone know*	1

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S5				132	*by definition*	1
				133	*as I mentioned*	1
				134	*are concerned about*	1
				135	*all it means*	1
				136	*according to*	1
M5S6	1	*we will see*	1	1	*notice that*	4
	2	*we should know*	1	2	*have to*	4
	3	*pay attention to*	1	3	*remember that*	3
	4	*keep this thing in your mind*	1	4	*one thing to notice*	3
				5	*the thing to notice*	2
				6	*bear in mind*	2
				7	*you need to remember*	1
				8	*remember [0.9] that*	1
				9	*one last note*	1
				10	*keep in mind*	1
				11	*be cautious of*	1
M5S7	1	*this is very very useful*	1	1	*have to*	12
	2	*this is important*	1	2	*I think*	10
	3	*it's really hard work*	1	3	*I mean*	4
	4	*it's really hard to*	1	4	*kind of*	2
	5	*it's interesting*	1	5	*there's no doubt*	1
	6	*it's creepy*	1	6	*that's very important*	1
	7	*I think*	1	7	*that's not the case*	1
	8	*have to*	1	8	*tend to be*	1
	9	*are the keys to*	1	9	*tend to*	1
				10	*tend not to*	1
				11	*sound like*	1
				12	*so to speak*	1
				13	*seem to*	1
				14	*one of the important parts*	1
				15	*of course*	1
				16	*making the difference*	1

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S7				17	*it's very straightforward*	1
				18	*it's very important*	1
				19	*it's very easy*	1
				20	*it's very difficult*	1
				21	*it's true*	1
				22	*it's obviously going to*	1
				23	*it's kind of*	1
				24	*it's important to*	1
				25	*it's easier to*	1
				26	*it's difficult to*	1
				27	*it seems that*	1
				28	*it seems obvious*	1
				29	*it is extremely useful*	1
				30	*I'm sure*	1
				31	*I'm not sure*	1
				32	*I'm afraid*	1
				33	*I thought*	1
				34	*I don't think*	1
				35	*I don't believe*	1
				36	*depends on*	1
				37	*be quite careful about*	1
				38	*a little bit*	1
				39	*a great deal*	1
M5S8	1	*let's review*	1	1	*what we're er assuming*	1
				2	*what I'm saying*	1
				3	*'ve gone through*	1
				4	*ends up with*	1

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S9	1	we can see	4	1	let's say	3
	2	do you understand	3	2	we're gonna	2
	3	we see	2	3	do you know	2
	4	let's go on	2	4	are there any	2
	5	if we want to	2	5	you'll notice	1
	6	do you know	2	6	you can imagine	1
	7	who can tell me	1	7	what would happen	1
	8	who can give me	1	8	what we're trying to	1
	9	we know	1	9	what we talked about	1
	10	we also can see	1	10	what sort of	1
	11	let's look at the picture	1	11	what happens	1
	12	let's look at	1	12	what do you think	1
	13	I'll show you	1	13	what about	1
	14	do you remember	1	14	let's think of	1
	15	do you know	1	15	I'm not gonna	1
	16	can you understand	1	16	I'm gonna	1
				17	if you think about it	1
				18	if we think about	1
				19	if there's not	1
				20	does that make sense	1
				21	does anyone have any idea	1
				22	does anybody know	1
				23	any questions	1
				24	any [2.1] questions	1
M5S10	1	what I want to say	2	1	I'm not going to	10
	2	I'm not going to	1	2	we're talking about	3
				3	what we're going to do	2
				4	we're gonna assume	2
				5	I'm not suggesting	2
				6	I'm not gonna	2
				7	I'm gonna	2

Table 6.8 MS FSs in *Move 5 (Cont.)*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M5S10				8	*I'm going to*	2
				9	*what we're not going to*	1
				10	*what we're going to get*	1
				11	*what we're doing*	1
				12	*what we are saying*	1
				13	*what I'm going to do*	1
				14	*what I'm going to*	1
				15	*what I'm doing*	1
				16	*we're gonna*	1
				17	*we're going to*	1
				18	*touch on*	1
				19	*the interesting thing about*	1
				20	*rather than*	1
				21	*let's not*	1
				22	*let's get it right*	1
				23	*in fact*	1
				24	*I'm talking about*	1
				25	*I'm not talking about*	1
				26	*I'm not going*	1
				27	*I'm gonna show*	1
				28	*I won't go through*	1
				29	*I wanna*	1
				30	*I should say*	1
				31	*amount to*	1

Table 6.8 presents the MS SFs (e.g., *be defined as*, *there's no doubt*, and *i'm not going to*) that CCL and CEL lecturers use in different steps of **M5 Elaborating The Topic** (See excerpt 148-159)).

148) <M5S1>Ok, geological resource can *be defined as* concentration of naturally occurring solid, liquid, or gas material in or on the Earth's crust in such forms and amount that economic extraction of a commodity from the concentration is currently or potentially feasible.</M5S1> (CCL03)

149) <M5S2>okay…what you *have to* remember is that we *are completely dependent on* plants [1.2] okay [0.4] we*'re not completely dependent on* animals we *are completely dependent on* plants [0.6] we either eat [0.3] them [0.3] or eat [0.5] the thing that ate them [0.9] okay [1.3] so whatever happens [0.3] …</M5S2> (CEL05)

150) <M5S3a>Because you may remember something, but I just today, I want to review, er, some basic concepts of hydraulic fracture, *refresh your memory*….</M5S3a> (CCL08)

151) <M5S3b>okay [0.4] you may ask where the leukocytes come from to do this test [0.4] *in the case of* a live related live donor [0.3] it's [0.3] perfectly obvious you take a little bit of blood [0.7] *in the case of* the cadaveric donor you use usually the spleen [0.4] as a sort of l-, source of leukocytes for this test…</M5S3b> (CEL07)

152) <M5S3c>now *with that in mind* , uh actually when we talk about the, the time required to execute the algorithm, we look at the, the most overriding factor and that is, actually, the number of comparisons. and that 's proportional to N-squared. …</M5S3c> (CEL12)

153) <M5S4> *Let's look at* the weathering. Weathering is, er, disintegration and decomposition of rocks in situ, of transportation involved. This a key words, in situ. That means the fragment rocks remain in the original site. They, they are not take away, right? …</M5S4> (CCL05)

154) <M5S5>running for the bus *for example* when you suddenly need a big increase in energy. um well, that is the, um, place where metabolic scope is highest. and for this reason , this is sometimes defined as , an optimum temperature.</M5S5> (CEL08)

155) <M5S6>… Ok, this, so when we applying the waterflooding curve, we should pay attention to dynamics of later development of the oilfields, and make regular analysis and the correction work. This is what we should note.</M5S6> (CCL09)

156) <M5S7>it is a very unpleasant death [1.5] you know [0.8] i mean it's a lingering death [0.6] you're ill for several months you have massive weight loss [0.4] people end up [0.4] bleeding often from all the orifices of the body [0.4] they've got all these opportunistic infections it's a very unpleasant way to die *there's no doubt* about that [1.0]</M5S7> (CEL06)

157) <M5S8>So *let's review* what we said , igneous rocks. What is igneous rocks? It is melted, it is formed when the melted rocks cools, and solidified. That's how the igneous rocks is formed and what is igneous rocks. And second is …</M5S8> (CCL02)

158) <M5S9>And this question is mean you should find equation of tangent line, and this tangent line passes through this point, yes? So we know that we have the formula y equal y zero equal the slope of tangent line multiple x minus x zero... ***Do you understand*** ? So if we rewrite in slope intercept form, just like this. This is the example one, example one.</M5S9> (CCL11)

159) <M5S10> i mean the interesting thing about this table [0.5] ***i'm not going to*** go through it in [0.6] great detail [0.4] but i want you to [0.4] focus your attention [1.4] on this column here [2.6]</M510> (CEL06)

Table 6.6 and table 6.8 show varied use of MS FSs between CCL and CEL lecturers in different steps of ***Move 5***. Chinese EMI lecturers used significantly more MS FSs in ***M5S3a Presenting background knowledge*** than their English counterparts, which shows great importance Chinese lecturers attached to background knowledge, as suggested in the over-reliance of the step ***M5S3a*** in section 4.3.

Table 6.6 and table 6.8 also show that Chinese lecturers used significantly less MS FSs in ***M5S5 Demonstrating the topic***, ***M5S6 Providing Caveats***, ***M5S7 Making Comments*** and ***M5S10 Pinning down and/or clarifying the topic*** than the native English lecturers. Meanwhile, much less MS FS types were found among the Chinese lecturers than the native English lecturers in these steps (33 vs 136 in ***M5S5***, 4 vs 11 in ***M5S6***, 9 vs 39 in ***M5S7***, and 2 VS 31 in ***M5S10***). On the other hand, findings in Section 4.3 reveal significantly less use of those steps of ***M5S6***, ***M5S7***, and ***M5S10*** by Chinese EMI lecturers than the native English lecturers, but no significant difference in step

M5S5 between them. All these seem to point to limited MS FS repertoire in these steps

of ***Move 5*** among the Chinese lecturers, particularly MS FSs in ***M5S5***.

Table 6.9 MS FSs in *Move 6*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M6S1	1	*according to*	2	1	*as we said*	4
	2	*let's see*	1	2	*remember that*	2
				3	*remind you of*	2
				4	*what I've just said*	1
				5	*looked at*	1
				6	*let's actually go through*	1
				7	*let me just recap*	1
				8	*is related [0.3] to*	1
				9	*heard of*	1
				10	*cast your mind back*	1
				11	*at the end of*	1
				12	*at the beginning of*	1
				13	*as I've already said*	1
				14	*as I was saying*	1
				15	*as I said*	1
				16	*as I pointed out*	1
M6S2	1	*we're going to talk about*	3	1	*I'm gonna*	3
	2	*later on*	3	2	*I'll talk about*	3
	3	*we're going to*	1	3	*we're going to see*	2
	4	*we'll talk about*	1	4	*what we're going to see*	1
	5	*we'll show later*	1	5	*we're gonna see*	1
	6	*we will talk about*	1	6	*we're gonna*	1
	7	*we are going to*	1	7	*we'll talk about*	1
				8	*later on*	1
				9	*in a second*	1
				10	*in a moment*	1
				11	*I'm gonna come back*	1
				12	*I'll come on to*	1

Table 6.9 MS FSs in *Move 6* (Cont.)

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M6S3	1	*has no relationship with*	2	1	*on the other hand*	2
	2	*depends on*	2	2	*in the case of*	2
	3	*as I have said*	2	3	*what I said earlier*	1
	4	*are parallel to*	2	4	*to remind you*	1
	5	*let's see*	1	5	*the disadvantage of*	1
	6	*let's compare with*	1	6	*the advantage of*	1
	7	*is linear to*	1	7	*make this distinction*	1
	8	*is depends on*	1	8	*is quite different from*	1
	9	*is depend on*	1	9	*is different from*	1
	10	*depend on*	1	10	*in the situation of*	1
	11	*as I have described*	1	11	*in contrast, to*	1
				12	*in contrast to*	1
				13	*in conjunction with*	1
				14	*has to do with*	1
				15	*different types of*	1
				16	*different from*	1
				17	*differences between*	1
				18	*difference between*	1
				19	*comparison between*	1

Table 6.9 presents the MS SFs (e.g., *remind you of*, *as I pointed out* and *let's compare with*) that CCL and CEL lecturers use in different steps of ***M6 Building Theme Network*** (See excerpt 160-162)).

160) <M6S1>Ok, and ***according to*** one hundred and thirty five waterflooding oilfields, including the seven carbonate reservoir, the professor Chen Yuanqian in China also got the empirical formula of the, of the, that kind of predict the reserve, dynamic reserve.</M6S1> (CCL09)

161) <M6S2>and this is ***what we're going to see*** [0.5] ***later on*** [0.7] er [0.2] in the

course [0.7] so we've got to [0.3] sacrifice something [0.2] if you like [0.7]

when we go to short fibres [0.6] we lose some of their potential [0.6] and [0.3]

when we go to [0.2] non-parallel fibres [0.3] we lose [0.2] something else

[1.3] but that [0.3] is [0.5] for [0.2] tomorrow [0.4]</M6S2> (CEL03)

162) <M6S3>…we can see that it ***has no relationship with*** the natural, er with the

geological reserve, right? It only ***depends on*** the viscosity ratio between oil

and water. This is the relationship or difference between the type A and type

B.</M6S3> (CCL09)

Table 6.6 and table 6.9 show significantly less use of MS FSs in ***M6S1 Referring***

to previous lecture/other source by the Chinese EMI lecturers than the native English

lecturers. It can be seen that Chinese EMI lecturers used only two types of MS FSs,

whereas the native English lecturers used as many as sixteen types of such FSs.

Structurally, CEL lecturers mainly used clausal and verb-based FSs to signal ***M6S1***.

Although no significant difference was found in the use of MS FSs in ***M6S2*** and ***M6S3***

between CCL and CEL lecturers, Chinese EMI lecturers used much less types of MS

FSs than the native English lecturers (7 vs 12 in ***M6S2*** and 11 vs 19 in ***M6S3***). All these

seem to suggest limited MS FS repertoire in ***Move 6*** among the Chinese EMI lecturers.

Table 6.10 presents the MS SFs (e.g. *I'm just curious* and *are you familiar with*)

that CEL lecturers use in ***M7 Making Aside*** (See excerpt 163)).

Table 6.10 MS FSs in *Move 7*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M7				1	*I'm just curious*	1
				2	*I can't imagine*	1
				3	*be familiar from with*	1
				4	*are you familiar with*	1

163) <M7>you can derive this from scratch … uh ***I'm just curious*** cuz this depends

on how many courses you took in statistics, how many of you have dealt

before with order statistics? By dealing with i mean, ***are you familiar with***

the topic? … okay.</M7> (CEL10)

Table 6.6 and table 6.10 reveal that no MS FSs was found in ***M7 Making Aside***

among the CCL lecturers, but four different types of MS FSs were used by the CEL

lecturers. Structurally, they are clausal and AP(C)-based FSs. The native English

lecturers used such FSs to masterfully discuss with students less relevant content, which

was generally more relaxing, thus beneficial for lively classroom atmosphere and good

teacher-student rapport.

Table 6.11 MS FSs in *Move 8*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M8				1	*remind you that*	1
				2	*I'm not going to*	1
				3	*if you're interested in*	1
				4	*if you want to*	1
				5	*I don't want to*	1
				6	*don't waste time doing*	1

Table 6.11 presents the MS SFs (e.g. *remind you that* and *if you're interested in*) that CEL lecturers use in *M8 Housekeeping* (See excerpt 164)).

164) <M8> *if you're interested in* them then both [0.3] er [0.7] two of the books that i i've referred to [0.2] Ward and Robinson [2.0] and Strangeways [2.0] er go into quite a lot of detail about snow measuring techniques …</M8> (CEL04)

Table 6.6 and table 6.11 show that no MS FSs was found in *M8* among the CCL lecturers, but six different types of MS FSs were used by the CEL lecturers. The MS FSs in *M8* are conducive to conveying teachers' recommendations and reminders.

Table 6.12 MS FSs in *Move 9*

Move		CCL			CEL	
	N	MS FS	tokens	N	MS FS	tokens
M9	1	*do you understand*	2	1	*are there any questions*	3
	2	*any question*	2	2	*any questions*	3
	3	*do you know*	1	3	*any other questions*	2
				4	*what's gonna happen*	1
				5	*what about*	1
				6	*I don't know*	1
				7	*any questions or comments*	1
				8	*any further questions*	1
				9	*answer some questions*	1

Table 6.12 presents the MS SFs (e.g. *any questions* and *any questions or comments*) that CCL and CEL lecturers use in *M9 Checking Comprehension & Consolidating Learning* (See excerpt 165)).

165) <M9>okay? *any questions*...? terrific.</M9> (CEL08)

Table 6.6 and table 6.12 show no significant difference in the use of MS FSs *in M9* between CCL and CEL lecturers. Nevertheless, the native English lecturers used much more types of MS FSs in *M9* than the Chinese EMI lecturers (9 vs. 3). Since this move generally serves to check on students' comprehension and to consolidate learning, most of the MS FSs include the key word 'question', or questions fragments to serve the purpose of the step.

6.4 Move signaling formulaic sequences in moves/steps of the Closing Phase

Table 6.13 presents the statistics of MS FSs in the moves/steps of the Closing Phase. It reveals that no significant difference was found in the use of MS FSs in any of the moves/steps in the Closing Phase between CCL and CEL lectures.

Table 6.13 Statistics of MS FSs in the Closing Phase

Move	MS FS		
	CCL (%)	CEL (%)	LL critical value G2[a]
M10	4	3	+1.42
M11S1	0	3	-1.72
M11S2	5	11	-0.10
M12S1	2	2	+0.38
M12S2	1	0	+2.10

Note. [a] The log-likelihood (LL) critical value G2 itself is always a positive number. However, indicators '+' and '-' are inserted to show overuse and underuse of corpus 1 (in this case CCL) relative to corpus 2 (in this case CEL). [*]$p < 0.05$, critical value G2= 3.84; [**]$p < 0.01$, critical value G2 = 6.63; [***]$p < 0.001$, critical value G2 = 10.83; [****]$p < 0.0001$, critical value G2 = 15.13.

Table 6.14 MS FSs in *Move 10*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M10	1	*we have talked about*	1	1	*we've talked about*	1
	2	*the key point is*	1	2	*in summing up*	1
	3	*let's review*	1	3	*finished up by*	1
	4	*let's make a summary*	1			

Table 6.14 presents the MS SFs (e.g. *let's make a summary* and *in summing up*) that CCL and CEL lecturers use in ***M10 Wrapping Up Lecture*** (See excerpt 166)).

166) <M10>Ok ***let's make a summary***. The mechanism and model, we have three

mechanism, one is what? Liquid film. The second is what? Liquid droplet.

The third one? Multi-droplet. …</M10> (CCL06)

Table 6.14 shows that both CCL and CEL lecturers used relatively few MS FSs to wrap up lectures. However, given the conventional status of ***M10*** in CCL (58%) and optional status in CEL (17%) respectively, it might be speculated that CCL lecturers may not be equipped with rich MS FS resources to wrap lectures while CEL lectures may not be so concerned with this lecture-wrapping move.

Table 6.15 MS FSs in *Move 11*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M11S1				1	*any questions*	1
				2	*anything else*	1
M11S2	1	*we'll learn*	1	1	*I'm gonna*	2
	2	*we'll discuss about*	1	2	*what we're going to do*	1
	3	*we will study about*	1	3	*we're going to*	1
	4	*we will learn*	1	4	*we'll talk about*	1
	5	*I will talk about*	1	5	*we'll see*	1
				6	*we'll look at*	1
				7	*we'll continue*	1
				8	*we are going to*	1
				9	*look forward to*	1
				10	*I'll start*	1

Table 6.15 presents the MS SFs (e.g. *what we're going to do* and *we'll talk about*) that CCL and CEL lecturers use in different steps of *M11 Cooling Down* (See excerpt 167-168)).

167) <M11S1>all right any ***anything else***</M11S1> (CEL02)

168) <M11S2>next week ***we'll talk about*** uh human populations and, allometry. <UNINTELLIGIBLE SPEECH></M11S2> (CEL11)

Table 6.15 reveals that while native English lecturers used MS FSs to explicitly check whether all homework-related reminders in step *M11S1 Housekeeping* were clearly received by students, Chinese EMI lecturers didn't use any MS FS to serve this function. Since step *M11S1* appeared in 42% (5/12) of lectures in both CCL and CEL, CCL lecturers may have used other linguistic devices to help achieve the rhetorical

function of this step.

Meanwhile, table 6.15 also indicates that both CCL and CEL lecturers mainly used pronoun-embedded MS FSs in the '**We/I + auxiliary + verb**' pattern to help signal the communicative function of *M11S2 Looking ahead*, except that CEL lecturers seem to have richer verb choices in MS FSs.

Table 6.16   MS FSs in *Move 12*

Move	CCL			CEL		
	N	MS FS	tokens	N	MS FS	tokens
M12S1	1	*the end of*	1	1	*the end of*	1
	2	*that's all*	1	2	*bring things to an end*	1
M12S2	1	*that's all*	1			

Table 6.16 presents the MS SFs (e.g. *that's all* and *bring things to an end*) that CCL and CEL lecturers use in different steps of *M12 Ending Lecture* (See excerpt 169-170)).

169) <M12S1>Now this is ***the end of*** our class.</M12S1> (CCL10)

170) <M12S2>Ok. ***That's all***. Thank you!</M12S2> (CCL09)

Table 6.16 shows that both CCL and CEL lecturers would spare the effort to explicitly put an end to the lecture through the use of MS FSs in *M12*, though no MS FS was used by CEL lecturers in *M12S2*.

6.5 Summary

This chapter reports the results and discussion of the functions of formulaic sequences in connection with rhetorical moves/steps where they are situated.

Statistical analyses show that the majority of FSs in the two corpora (CCL: 85% and CEL: 95%) are directly or indirectly connected with rhetorical moves/steps. In terms of actual occurrences, the Chinese lecturers used significantly less MS FSs and MB FSs than the native English lecturers, but significantly more MN FSs than the latter. The study reported only the use of MS FSs in CCL and CEL.

In the Opening Phase, Chinese lecturers used significantly less MS FSs in *M2S4*, but significantly more MS FSs in *M3S2* than the native English lecturers. In the Theme Network Building Phase, Chinese lecturers used significantly less MS FSs in *M5S5*, *M5S6*, *M5S7*, *M5S10*, *M6S1*, and *M8*, but significantly more MS FSs in *M5S3a* than their native English counterparts. The Closing Phase doesn't show any significant differences between CCL and CEL lecturers in terms of MS FS use in any move/step.

The major differences in MS FS use between CCL and CEL lecturers seem to suggest that Chinese lecturers might have a limited FS repertoire, MS FS inventories in particular.

CHAPTER 7

CONCLUSION

This chapter concludes the research. First, it presents a summary of the major findings of the research, addressing each of the research questions. Then, it examines the implications of the research. Finally, it discusses a few limitations involved and accordingly makes some suggestions for further research.

7.1 Summary of the Findings

Aiming at exploring both the macro-structure and the micro-linguistic features of university EMI lectures across cultures, the present research made a contrastive analysis of the rhetorical move structure and formulaic sequences in EMI lectures given by native and non-native English lecturers. Specifically, the research intends to: 1) identify and compare the rhetorical move structure of university EMI lectures given by native and non-native English lecturers; 2) identify and compare formulaic sequences used in university EMI lectures given by native and non-native English lecturers; 3) examine formulaic sequences in relation to the moves/steps to which they belong. The major findings are summarized in the following sections.

7.1.1 Rhetorical move structure

Twelve moves with various steps were identified in CCL and CEL. Log likely-hood analysis suggests slightly more moves and steps used by Chinese lecturers than by English lecturers, indicating that Chinese EMI lecturers are aware of rhetorical structure of the lecture genre, at least, as well as their native English counterparts, which was corroborated by the interview data.

All moves/steps (except that there is no *M2S3* in CCL and *M2S1* CEL) appear in both CCL and CEL. Most moves/steps in the two corpora share similar conventionality status, though there are some differences throughout the lectures. It seems to suggest that CCL and CEL share similar rhetorical move structure.

However, different lecturing preferences can still be found in the use of individual moves and steps between CCL and CEL lecturers, with the major ones centering upon the Theme Network Building Phase.

Chinese EMI lecturers seem to pay attention to ensuring the completeness of lectures, as revealed in significantly more use of moves/steps in the Opening (e.g., *M1 Getting Started*, *M2S1 Leading in* and *M3S2 Providing lecture scope*) and Closing Phase (e.g., *M10 Wrapping Up Lecture*). What's more, Chinese EMI lecturers also used significantly more moves/steps that involve established knowledge/information, e.g., *M4 Introducing The Topic, M5S1 Explaining terms*, *M5S3a Presenting background knowledge* and *M5S4 Expounding rationale*.

On the other hand, native English lecturers seem to be more comfortable with

and better at making connections between real life and knowledge as well as among knowledge nodes, as indicated by significantly more use of *M5S3b Contextualizing real-word experience* and *M6S1 Referring to previous lecture/other source.* Besides, native English lecturers also appear to be more willing to share with students their own critical understanding of issues under discussion, as suggested by significantly more use of *M5S6 Providing caveats* and *M5S7 Making comments*. Meanwhile, native English lecturers are more comfortable with and more experienced in attempting various strategies to promote classroom teaching effectiveness, as reflected in significantly more use of *M2S4 Housekeeping*, *M5S10 Pinning down and/or clarifying the topic*, *M8 Housekeeping*, and *M9 Checking Comprehension & Consolidating Learning*.

In addition, an ELF perspective of CCL lectures didn't find any move simplification. However, move hybridity in the form of move cycles did occur in CCL lectures. Five move cycles were found unique to CCL lectures, including *M5S4-M5S4*, *M5S5-M5S4*, *M6S3-M5S3a*, *M1-M3S1*, and *M3S1-M3S2*.

7.1.2 Formulaic sequences

A total of 1615 and 4354 formulaic sequences were identified in CCL and CEL lectures, respectively. The log likely-hood comparison revealed significantly less use of FSs in terms of both types and tokens by Chinese EMI lecturers than by their native English counterparts. It is speculated that Chinese EMI lecturers might possess a more limited repertoire of FSs than the native English lecturers.

On the basis of Biber et al (1999) and Wang (2017), the present study classified the FSs into eight different structural categories, including NP fragment, PP fragment, AP fragment as modifier, AP fragment as complement, Clausal fragment, VP fragment, AdP fragment, and ConjP fragment. In term of token frequencies, lecturers mainly relied on NP-, PP-, clausal, VP-, and ConjP-based FSs. Both CCL and CEL exhibit a mixed feature of 'oral' and 'literate' characteristics.

An ELF perspective of analysis identified 34 non-standard FS types and 63 non-standard FS tokens by the CCL lecturers, accounting for 6% of the total FS types and 4% of the total FS tokens respectively. These non-standard FSs were summarized into nine different categories in terms of their lexico-grammatical features. Nonetheless, in congruence with Björkman (2008b), none of them seem to cause any disturbance or breakdown in communication.

7.1.3 Move-formulaicity connections

The present study also examined the functions of FSs in connection with the rhetorical moves/steps. Statistical comparison suggests that the majority of FSs in the two corpora (CCL: 85% and CEL: 95%) are directly or indirectly connected with rhetorical moves/steps. In terms of actual occurrences, the Chinese lecturers used significantly less MS FSs and MB FSs than the native English lecturers, but significantly more MN FSs than the latter.

In the Opening Phase, Chinese EMI lecturers used significantly less MS FSs in *M2S4*, but significantly more MS FSs in *M3S2* than the native English lecturers. In

the Theme Network Building Phase, Chinese EMI lecturers used significantly less MS FSs in *M5S5*, *M5S6*, *M5S7*, *M5S10*, *M6S1*, and *M8*, but significantly more MS FSs in *M5S3a* than their native English counterparts. The Closing Phase doesn't show any significant difference between CCL and CEL lecturers in terms of MS FS use in all the moves/steps.

The major differences in MS FS use between lecturers of the two corpora seem to suggest that Chinese lecturers might have a limited FS repertoire, MS FSs in particular.

7.2 Implications of the research

As stated in Section 1.5, the present study has the potential to promote research in various areas of applied linguistics and ELT, particularly in classroom discourse studies, genre studies, and EMI pedagogy.

7.2.1 Implications for classroom discourse studies

As stated earlier, studies on EMI lectures in China are mainly theoretical and policy discussions (Hu, Li & Lei, 2014), lacking empirical research on classroom discourse. The present research attempts to fill the gap through the investigation of rhetorical move structure and formulaic language in real EMI lectures collected from a Chinese university as well as real academic lectures from MICASE and the BASE corpus. The research expects to present a real linguistic and discoursal landscape of what transpires in university EMI lectures in China.

7.2.2 Implications for genre studies

Due in part to the complexities involved in classroom discourse, genre studies have been generally focused on written data. Though the availability of established spoken corpora (e.g., MICASE, BASE, VOICE, EFLA) has greatly boosted research in academic lectures by native and non-native English lecturers, most of them have been interested in either specific linguistic elements, e.g., pronouns, questions, metadiscourse, lexical bundles, or lecture sections, e.g., lecture introductions, lecture closings.

Other than the IRF classroom exchange patterns which might be a bit incompatible with modern classroom discourse, the present research investigated the macro-structure of science-oriented lectures under the Swalesian ESP framework. The rhetorical move structure identified in the present study may contribute to advancing genre studies in strengthening the legitimate status of EMI lectures as a well-established genre.

In addition, a contrastive analysis of rhetorical move structure between native and non-native English lecturers expects to shed light on unveiling potential differences in terms of lecture schemata internalized by lecturers across cultures.

It is also worth mentioning the academic value of situating formulaic sequences in rhetorical move structure. Though some researchers have attempted to link formulaic sequences (lexical bundles/clusters) to their rhetorical moves in research articles or research article sections, they all focused on written data. In addition, these

studies adopted a bottom-up methodology, starting from the identification of lexical bundles/clusters and analyzing them in the extended contexts to establish the relationship between the two categories. This invites the flaw of disregarding those textual segments where no lexical bundles/clusters appear, thus weakening the validity of the results to some extent. Meanwhile, the methodologies in previous studies also mistakenly took textual proximity for bundle-move connections. Nevertheless, the present research attempts to overcome these flaws and resorts to intuitive judgment of the move-formulaicity connections according to systematically designed criteria, which deal directly with the connections *per se* instead of circumventing the topic by investigating peripheral elements. Therefore, it is more to the point.

7.2.3 Implications for EMI pedagogy

Different from previous studies on lecture rhetorical move structure (e.g., Young, 1994; Dudley-Evans, 1994; Thompson, 1994; Lee, 2009) that focused on improving students' lecture comprehension, this study is more concerned with difficulties of non-native English lecturers in delivering coherent disciplinary content. Though not intended as a straitjacket of a genre model, the rhetorical move structure identified from lectures of MICASE and BASE and Chinese EMI classrooms undoubtedly will have significant pedagogical implications for EMI lecturers. In the practical sense, EMI lecturers in China are sure to benefit from the comparisons, especially considering the perceived rarity of successful EMI experience in Chinese universities.

Meanwhile, the present study also generates a list of formulaic sequences, which might be, to a great extent, able to lessen the processing effort for teachers and students alike in the lecturing process. The knowledge of these formulaic sequences promises to help L2 lecturers to deliver the lesson in a more spontaneous and freewheeling manner.

Finally, the rhetorical move structure labelled with move signaling formulaic sequences should be of immense value for educational practitioners in EFL and ELF contexts. The rhetorical move structure, together with corresponding move signaling formulaic sequences, of EMI lectures by native English lecturers is presented in Table 7.1. L2 lecturers, especially EMI lecturers, would always benefit from it whenever the needs arise.

Table 7.1 **Rhetorical moves and formulaic sequences**

Moves/steps	Move signaling formulaic sequences (N) [a]
M1 Getting Started	*shall we start, Let's start, why don't we get started* (3)
M2 Warming Up	
M2S1 Leading in	*N/A* [b]
M2S2 Recalling previous lecture(s)	*what I said, we've been talking about, we talked about, had looked at, bring you back to, at the end of* (6)
M2S3 Looking ahead	*you're gonna see, I'm going to* (2)
M2S4 Housekeeping	*you'll see, what I wanted to say, we'll come back to, I'm not going to, I'm just gonna, I'm gonna, I'm going to, I'd like to, feel free to ask, are there any questions, any questions, any comments* (12)

Table 7.1 **Rhetorical moves and formulaic sequences (Cont.)**

Moves/steps	Move signaling formulaic sequences (N) [a]
M3 Setting Up Lecture Agenda	
M3S1 Announcing/ clarifying lecture theme	*we're going to, talk about, what I'm going to, what I want to talk about, what I want to do, we're gonna, we move on, start off by, let's actually get on with, I'm gonna, I'm going to, I'd like to, be talking about* (13)
M3S2 Providing lecture scope	*go through, what I'd like to do, what I want to cover, we're gonna say, we're gonna assume, I'm going to talk, go on to, at the end* (8)
M4 Introducing The Topic	*we're going to, look at, talk about, what about, think about, let's look at, I'm gonna, worry about, we move on to, touch on, the question is, let's go on, I wanna, what we're going to, what we're doing, what we'd like to do, what we wanna do, what I'm going to do, what I'd like to do, we're gonna see, we're gonna, we're going to look at, we'll get on, we go on to, there's another thing, there's another issue, the question was, the next thing, take a look at, one of the questions, one of the problems, move along to, make the next argument, let's talk about, let's take a look at, let's start looking at, let's start with, let's look, let's have a look, let's go to, let's go on to, let's move on, let's go on, let's briefly look at, let us think, lemme point out, lemme get back to, I'm, gonna, I'm just gonna, I'm going to, I'll talk about, I'll show you, if we're going to, I want to talk about, going back to, go through, go about trying, deal with, come across* (59)

Table 7.1 **Rhetorical moves and formulaic sequences (Cont.)**

Moves/steps	Move signaling formulaic sequences (N) [a]
M5 Elaborating The Topic	
M5S1 Explaining terms	*we're gonna assume, let me introduce, it means, by definition* (4)
M5S2 Highlighting importance	*are completely dependent on, 're not completely dependent on, it's important, have to* (4)
M5S3a Presenting background knowledge	*if you look at, the case of, 're well familiar with, it has been said, in many cases, if you're looking at, if you think about it, if we look at, heard of* (9)
M5S3b Contextualizing real-word experience	*in the case of, you've probably seen, what about, under those circumstances, let's think of, in this case, in the morning, if you look at, be familiar with* (9)
M5S3c Setting up premise	*with that in mind, when we talk about, it is the case that, if you remember* (4)
M5S4 Expounding rationale	*in terms of, in other words, for example, in fact, I think, for instance, the idea of, such as, in the case of, in practice, I mean, what we're going to see, what we are assuming, what I'm trying to, what I mean, we're gonna assume, the question is, the problem is,* *the point that, the point is, the idea was, the fact that, the assumption is, that means, one of the things, make the assumption, look at, let's just think about, let's assume, it means that, is defined as, in the context, in that situation, in some cases, in principle, if you're looking at, if we look at, I should say, for that matter, for example, focus on, as a principle, are referred to* (43)

Table 7.1 **Rhetorical moves and formulaic sequences (Cont.)**

Moves/steps	Move signaling formulaic sequences (N) [a]
M5S5 Demonstrating the topic	*for example, and so on, in fact, in other words, I mean, I think, in terms of, you can see, we're gonna, look at, let's look at, I'm gonna, what happens, think of, that's why, in the case of, as you can see, what you can see, we'll look at, that means, let's say, it's estimated that, you'll see, with respect to, what we're doing, we're looking at, we can see, we are going to, think about, the definition of, refer to this as, looking at, let's go back to, let us suppose, I'm gonna show you, I'm going to, if you imagine, I should say, I believe, for instance, concentrate on, you'll notice, you would say, you know, when we talked about, what's happened, what we're saying, what we're really doing, what we have to do, what we expect to see, what we can see, what we are going to be doing, what this meant, what this means, what it means, what I'm gonna, what I'm doing, what I mean, what am I talking about, we're thinking of, we're gonna use, we're about to, we expect to see, we can see that, we can also see, this issue of, the trouble is, the reason for that, the question is, the problem of, the problem is, the key feature to, the issue of, the idea of, the first thing to, the first point to note, that is why, such as, start off with, referring to, refer to as, refer to, one thing to mention, one of the things, let's think about, let's then see, let's take a look at, let's take a look again at, let's start off with, let's see, let's just take a quick look at, let's just say, let's have a quick look, let's have a look at, let's go back, let's also suppose, let's look at, let me see, let me get to that, lemme go back to, it's estimated, it turns out that, it stacks up to, it means, is looking at, is concerned about, in virtually all case, in this case, in theory, in that case, in terms of,*

Table 7.1 **Rhetorical moves and formulaic sequences (Cont.)**

Moves/steps	Move signaling formulaic sequences (N) [a]
M5S5 Demonstrating the topic	*in some cases, in principle, in particular, in many cases, in effect, in all cases, I'm looking for, if you think of, if you think about it, if you look at, if you look, if we try to, if we look at, if we imagine, if we assume, I would like you to, I suppose, here you see, focused on, does anyone know, by definition, as I mentioned, are concerned about, all it means, according to* (136)
M5S6 Providing caveats	*notice that, have to, remember that, one thing to notice, the thing to notice, bear in mind, you need to remember, remember that, one last note, keep in mind, be cautious of* (11)
M5S7 Making comments	*have to, I think, I mean, kind of, there's no doubt, that's very important, that's not the case, tend to be, tend to, tend not to, sound like, so to speak, seem to, one of the important parts, of course, making the difference, it's very straightforward, it's very important, it's very easy, it's very difficult, it's true, it's obviously going to, it's kind of, it's important to, it's easier to, it's difficult to, it seems that, it seems obvious, it is extremely useful, I'm sure, I'm not sure, I'm afraid, I thought, I don't think, I don't believe, depends on, be quite careful about, a little bit, a great deal* (39)
M5S8 Summing up the topic	*what we're er assuming, what I'm saying, 've gone through, ends up with* (4)
M5S9 Initiating co-building lecture	*what we talked about, what sort of, what happens, what do you think, what about, let's think of, I'm not gonna, I'm gonna, if you think about it, if we think about, if there's not, does that make sense, does anyone have any idea, does anybody know, any questions, any questions* (24)

Table 7.1 **Rhetorical moves and formulaic sequences (Cont.)**

Moves/steps	Move signaling formulaic sequences (N) [a]
M5S10 Pinning down and/or clarifying the topic	*I'm not going to, we're talking about, what we're going to do, we're gonna assume, I'm not suggesting, I'm not gonna, I'm gonna, I'm going to, what we're not going to, what we're going to get, what we're doing, what we are saying, what I'm going to do, what I'm going to, what I'm doing, we're gonna, we're going to, touch on, the interesting thing about, rather than, let's not, let's get it right, in fact, I'm talking about, I'm not talking about, I'm not going, I'm gonna show, I won't go through, I wanna, I should say, amount to (31)*
M6 Building Theme Network	
M6S1 Referring to previous lecture/other Source	*as we said, remember that, remind you of, what I've just said, looked at, let's actually go through, let me just recap, is related [0.3] to, heard of, cast your mind back, at the end of, at the beginning of, as I've already said, as I was saying, as I said, as I pointed out (16)*
M6S2 Specifying subsequent/future content	*I'm gonna, I'll talk about, we're going to see, what we're going to see, we're gonna see, we're gonna, we'll talk about, later on, in a second, in a moment, I'm gonna come back, I'll come on to (12)*
M6S3 Connecting/ Comparing topics	*on the other hand, in the case of, what I said earlier, to remind you, the disadvantage of, the advantage of, make this distinction, is quite different from, is different from, in the situation of, in contrast, to, in contrast to, in conjunction with, has to do with, different types of, different from, differences between, difference between, comparison between (19)*

Table 7.1 **Rhetorical moves and formulaic sequences (Cont.)**

Moves/steps	Move signaling formulaic sequences (N) [a]
M7 Making Aside	*I'm just curious, I can't imagine, be familiar from with, are you familiar with* (4)
M8 Housekeeping	*remind you that, I'm not going to, if you're interested in, if you want to, I don't want to, don't waste time doing* (6)
M9 Checking Comprehension & Consolidating Learning	*are there any questions, any questions, any other questions, what's gonna happen, what about, I don't know, any questions or comments, any further questions, answer some questions* (9)
M10 Wrapping Up Lecture	*we've talked about, in summing up, finished up by* (3)
M11 Cooling Down	
M11S1 Housekeeping	*any questions, any anything else* (2)
M11S2 Looking ahead	*I'm gonna, what we're going to do, we're going to, we'll talk about, we'll see, we'll look at, we'll continue, we are going to, look forward to, I'll start* (10)
M12 Ending Lecture	
M12S1 Dismissing class	*the end of, bring things to an end* (2)
M12S2 Farewell	*N/A*

Note. [a] *The number in the bracket refers to the types of formulaic sequences identified in the move/step.* [b] *N/A indicates that no formulaic sequence was identified in that specific move/step.*

7.3 Limitations and further research

The findings of the research promise to advance our understanding of the potential differences between EMI lectures given by native and non-native English lecturers in terms of both the macro-structure and the micro-linguistic elements. However, the qualitative nature of the present study also lends itself to potential limitations:

1) The corpora used for the research are relatively small. Only 24 lectures are used

With these limitations in mind, I strongly recommend that due attention be paid to the following aspects in further studies.

Firstly, larger corpora that include lectures from different disciplines and in different instruction languages are needed to test previous findings and to enrich genre theories. Disciplinary variations have proven to be essential in written discourse, but relatively little is known about the spoken genre. Meanwhile, it would be interesting to investigate lectures taught in different languages by the same lecturers, which may show whether and to what degree teaching in a different language other than their mother language would affect the lecture organization and formulaic language they use.

Secondly, it's advisable that further research should take demographic background of lectures and cultural factors into consideration to give thicker descriptions of EMI lectures, so as to corroborate findings of previous research.

Thirdly, follow-up interviews are recommended to be conducted within a relatively short time after the lecture so that lecturers' real intentions and understandings could be timely interviewed and recorded.

Fourthly, international cooperation might be conducive to contrastive study of EMI lectures across cultures. A well-coordinated international project may help enhance the synchronicity and contemporariness of data collection so the results could be of more theoretical and pedagogical value.
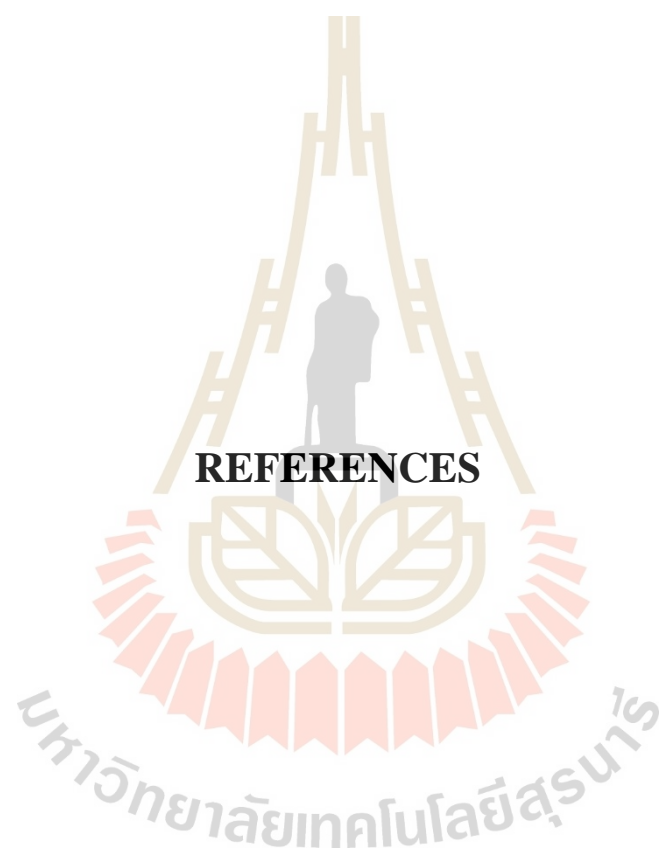
REFERENCES

REFERENCES

Ädel, A. (2010). Just to give you kind of a map of where we are going: A taxonomy of metadiscourse in spoken and written academic English. *Nordic Journal of English Studies*, 9(2), 69-97.

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English. *English for Specific Purposes*, 31, 81-92

Aguilar, M., & Rodríguez, R. (2012). Lecturer and student perceptions on CLIL at a Spanish university. *International Journal of Bilingual Education and Bilingualism*, 15, 183-197.

Airey, J. (2016). CLIL and EAP (Content and Language Integrated Learning and English for Academic Purposes). In K. Hyland & P. Shaw (Eds.), *Routledge handbook of English for academic purposes* (pp. 71-83). Abingdon: Routledge.

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, 17, 51-62.

Allan, R. (2016). Lexical bundles in graded readers: To what extent does language restriction affect lexical patterning?. *System*, 59, 61-72.

Allison, D., & Tauroza, S. (1995). The effect of discourse organization on lecture comprehension. *English for Specific Purposes*, 14(2), 157-173.

Altenberg, B. (1990). Speech as linear composition. In Caie, G. Haastrup, K., Jakobsen, A.L., Nielsen, J.E., Sevaldsen, J., Specht, H. and Zettersten, A. (Eds.), *Proceedings from the Fourth Nordic Conference for English Studies*, Vol. 1. Department of English, University of Copenhagen, pp. 133-143.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press, pp. 101- 22.

Amnuai, W., & Wannaruk, A. (2013). A move-based analysis of the conclusion sections of research articles published in international and Thai journals. 3L: Language, Linguistics and Literature, *The Southeast Asian Journal of English Language Studies*, 19(2), 53-63.

Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes*, 31, 58-71.

Anthony, L. (2018). AntConc (Version 3.5.0) [Computer Software]. Tokyo, Japan: Waseda University. Accessed from http://www.laurenceanthony.net/ November 13, 2018.

Aston, G. (2001). Text categories and corpus users: a response to David Lee. *Language learning and technology*, 5(3), 73-76.

Atkinson, D. (2004). Contrasting rhetorics/contrasting cultures: Why contrastive rhetoric needs a better conceptualization of culture. *Journal of English for Academic Purposes*, *3*(4), 277-289.

Bakhtin, M. (1986). The problem of speech genres. In C. Emerson & M. Holquist (Eds.), *Speech genres and other late essays* (pp. 66-102). Austin: University of Texas.

Ball, P., & Lindsay, C. (2013). Language demands and support for English-medium instruction in tertiary education: Learning from a specific context. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 44-61). Bristol, UK: Multilingual Matters.

Barr, P. (1990). The role of discourse intonation in lecture comprehension. In M. Hewings (Ed.). *Papers in discourse intonation* (Vol. 16, pp. 5-21). Birmingham, UK: English Language Research, University of Birmingham, Discourse analysis monograph.

Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in dentistry and disciplinary variation. *Journal of English for Academic Purposes*, *11*(2), 134-144.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press.

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 79-101). London: Taylor & Francis.

Belcher, D. (2006). English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly*, 40(1), 133-156.

Belica, C. (1996). Analysis of temporal changes in corpora. *International journal of corpus linguistics*, 1(1), 61-73.

Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, 69, 65-78.

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

Bhatia, V. K. (1997). Genre-mixing in academic introductions. *English for specific purposes*, *16*(3), 181-195.

Bhatia, V. K. (1998). Generic patterns in fundraising discourse. *New Directions for Philanthropic Fundraising*, 22, 95-110.

Bhatia, V. K. (2004). Worlds of written discourse: A genre-based view. New York: A&C Black.

Biber, D. & F. Barbieri (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam; Philadelphia: John Benjamins.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275-311.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure* (Vol. 28). Amsterdam: John Benjamins.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371-405.

Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.

Björkman, B. (2008a). English as the lingua franca of engineering: The morphosyntax of academic speech events. *Nordic Journal of English Studies*, 7(3), 103-122.

Björkman, B. (2008b). 'So where we are?' Spoken lingua franca English at a technical university in Sweden. *English today*, 24(2), 35-41.

Björkman, B. (2010). So you think you can ELF: English as a lingua franca as the medium of instruction. *HERMES-Journal of Language and Communication in Business*, 23(45), 77-96.

Björkman, B. (2011). Pragmatic strategies in English as an academic lingua franca: Ways of achieving communicative effectiveness?. *Journal of Pragmatics*, 43(4), 950-964.

Bolton, K., & Botha, W. (2015). English in China's universities: Past and present. *World Englishes*, 34(2), 190-210.

Bowles, H., & Cogo, A. (Eds.). (2015). *International perspectives on English as a lingua franca: pedagogical insights*. Hampshire: Palgrave Macmillan.

Breiteneder, A. (2005). The naturalness of English as a European lingua franca: the case of the 'third person -s'. *Vienna English Working Papers*, 14, 3-26.

Breiteneder, A. (2009). English as a lingua franca in Europe: an empirical perspective. *World Englishes*, 28 (2), 256-269.

British Council. (2013) *British Council regional policy dialogue 2: The role of English in higher education: Issues, policy and practice*. https://www.britishcouncil.org.br/sites/default/files/regionalpolicydialogue_emi.pdf. Accessed May 13, 2018.

Brown, G., & Yule, G. (1983a). *Discourse analysis*. Cambridge: Cambridge university press.

Brown, G., & Yule, G. (1983b). *Teaching the spoken language* (Vol. 2). Cambridge: Cambridge University Press.

Buerki (2016) Formulaic sequences: a drop in the ocean of constructions or something more significant? *. *European Journal of English Studies*, 20:1, 15-34.

Bunton, D. (2002). Generic moves in PhD thesis Introductions. In J. Flowerdew (Ed.), *Academic discourse* (pp. 57-75). London: Longman.

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30, 38-52.

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5(5), 31-64.

Cai, (Luna) Jing. (2016). An exploratory study on an integrated genre-based approach for the instruction of academic lexical phrases. *Journal of English for Academic Purposes*, 24, 58-74.

Carey, R. (2013). On the other side: Formulaic organizing chunks in spoken and written academic ELF. *Journal of English as a Lingua Franca*, 2(2), 207-228.

Chang, Y. Y. (2012). The use of questions by professors in lectures given in English: Influences of disciplinary cultures. *English for Specific Purposes*, 31(2), 103-116.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge: Cambridge University Press.

Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14, 30-49.

Cheng, S. W. (2012). "That's it for today": Academic lecture closings and the impact of class size. *English for Specific Purposes*, 31(4), 234-248.

Cho, H., & Yoon, H. (2013). A corpus-assisted comparative genre analysis of corporate earnings calls between Korean and native-English speakers. *English for Specific Purposes*, *32*(3), 170-185.

Clayman, S. E. (1993). Reformulating the question: A device for answering/not answering questions in news interviews and press conferences. *Text-Interdisciplinary Journal for the Study of Discourse*, 13(2), 159-188.

Cogo, A. (2009). Accommodating difference in ELF conversations: a study of Pragmatic strategies. In Mauranen, A., Ranta, E. (Eds.), *English as a Lingua Franca: Studies and Findings*. Newcastle: Cambridge Scholars Publishing, pp. 254-270.

Cogo, A. (2012). English as a lingua franca: Concepts, use, and implications. *ELT journal,* 66(1), 97-105.

Cogo, A., & Dewey, M. (2006). Efficiency in ELF communication: From pragmatic motives to lexico-grammatical innovation. *Nordic Journal of English Studies*, 5(2), 59-93.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29, 72-89.

Connor, U. (2000). Variation in rhetorical moves in grant proposals of US humanists and scientists. *Tex*t, 20(1), 1-28.

Connor, U. (2004). Intercultural rhetoric research: Beyond texts. *Journal of English for Academic Purposes*, *3*(4), 291-304.

Connor, U., & Mauranen, A. (1999). Linguistic analysis of grant proposals: European Union research grants. *English for Specific Purposes*, 18(1), 47-62.

Connor, U., Precht, K., & Upton, T. (2002). In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 175-194). Amsterdam: John Benjamins.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for academic purposes*, 12(1), 33-43.

Costa, F., & Coleman, J. A. (2013). A survey of English-medium instruction in Italian higher education. *International Journal of Bilingual Education and Bilingualism*, 16, 3-19.

Cotos, E., Huffman, S., & Link, S. (2017). A move/step model for methods sections: Demonstrating Rigour and Credibility. *English for Specific Purposes*, 46, 90-106.

Cots, J. M. (2013). Introducing English-medium instruction at the University of Lleida, Spain: Intervention, beliefs and practices. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 106-128). Bristol, UK: Multilingual Matters.

Coyle, D., P. Hood, and D. Marsh. (2010). Content and Language Integrated Learning. Cambridge: Cambridge University Press.

Crawford Camiciottoli, B. (2004). Interactive discourse structuring in L2 guest lectures: Some insights from a comparative corpus-based study. *Journal of English for Academic Purposes*, 3(1), 39-54.

Crawford Camiciottoli, B. (2008). Interaction in academic lectures vs. written text materials: The case of questions. *Journal of Pragmatics*, 40(7), 1216-1231.

Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication*, 10(1), 39-71.

Crossley, S., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *International Review of Applied Linguistics in Language Teaching*, 49(1), 1-26.

Crystal, D. (2004). The past, present, and future of World English. In Gardt & Huppauf (Eds.) *Globalization and the future of German: With a select bibliography* (pp. 27-45). Berlin: Mouton de Gruyter.

Csomay, E. (2005). Linguistic variation within university classroom talk: A corpus-based perspective, *Linguistics and Education*, 15, 243-274.

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied linguistics*, 34(3), 369-388.

Csomay, E., Jones, K., & Keck, C. (2007). Introduction to the identification and analysis of vocabulary-based discourse units. In Biber, Connor & Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure. Studies in corpus linguistics*, 28, 155-173.

Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms* (Vol. 20). Amsterdam: John Benjamins Publishing.

Dalton-Puffer, C., Nikula, T., & Smit, U. (Eds.). (2010). *Language use and language learning in CLIL classrooms* (Vol. 7). Amsterdam: John Benjamins Publishing.

De Beaugrande, R. (2001). Large corpora, small corpora, and the learning of language. In M. Ghadessy, A. Henry and R. Roseberry (Eds.), *Small Corpus Studies and ELT*, 3-28. Amsterdam: John Benjamins.

Dearden, J. (2014). *English as a medium of instruction-a growing global phenomenon. British Council*. https://www.britishcouncil.org/sites/default/files/e484_emi_-_cover_option _3_final_web.pdf. Accessed May 17, 2018.

DeCarrico, J., & Nattinger, J. R. (1988). Lexical phrases for the comprehension of academic lectures. *English for Specific Purposes*, 7, 91-102.

Deroey, K. L., & Taverniers, M. (2012). Just remember this: Lexicogrammatical relevance markers in lectures. *English for Specific Purposes*, 31(4), 221-233.

Devitt, A. J. (1991). Intertextuality in tax accounting: Generic, referential, and functional. In C. Bazerman & J. Paradis (Eds.), *Textual dynamics of the professions* (pp. 336-357). Madison: University of Wisconsin Press.

Devitt, A. J. (1993). Generalizing about genre: New conceptions of an old concept. *College Composition and Communication*, 44, 573-586

Dewey, M. (2007). *English as a Lingua Franca: An Empirical Study of Innovation in Lexis and Grammar*. Unpublished PhD dissertation. King's College, London.

Dias, P. (1994). Initiating students into the genres of discipline-based reading and writing. In A. Freedman & P. Medway (Eds.), *Learning and teaching genre* (pp. 193-206). Portsmouth, NH: Boynton/Cook.

Doiz, A., Lasagabaster, D., & Sierra, J. M. (2013). English as L3 at a bilingula university in the Basque County, Spain. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 84-105). Bristol, UK: Multilingual Matters.

Dong, J., & Buckingham, L. (2018). The collocation networks of stance phrases. *Journal of English for Academic Purposes*, 36, 119-131.

Dos Santos, M. B. (1996). The textual organization of research paper abstracts in applied linguistics. *Text-Interdisciplinary Journal for the Study of Discourse*, 16(4), 481-500.

Dubois, B. L. (1980). Genre and structure of biomedical speeches. *Forum Linguisticum*, 5, 140-169

Dudley-Evans, T. (1986). Genre analysis: An investigation of the introduction and discussion sections of MSc dissertations. In M. Coulthard (Ed.), *Talking about text* (pp. 128-145). Birmingham, UK: English Language Research, Birmingham University.

Dudley-Evans, T. (1994). Variations in the discourse patterns favoured by different disciplines and their pedagogical implications. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 146-158). Cambridge: Cambridge University Press.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28(3), 157-169.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165-193.

Durrant, P., & Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, 30(1), 58-72.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47, 157-177.

Eckert, P. (2000). *Linguistic variation as social practice*. Malden, Mass: Blackwell.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology*, 49A, 234-250.

Erling, E. J., & Bartlett, T. (2006). Making English their own: The use of ELF among students of English at the FUB. *Nordic Journal of English Studies*, 5(2), 9-40.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. Text-Interdisciplinary Journal for the Study of Discourse, 20(1), 29-62.

Feng, H. (2008). A genre-based study of research grant proposals in China. In U. Connor, E. Nagelhout & W. Rozycki (Eds.). *Contrastive Rhetoric: Reaching to intercultural rhetoric* (pp. 63-86). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Firth, A. (1996). The discursive accomplishment of normality: On 'lingua franca' English and conversation analysis. *Journal of pragmatics*, 26(2), 237-259.

Firth, J.R. (1957) *Papers in linguistics: 1934-51*. London: Oxford University Press.

Flowerdew, J. (1993). An educational, or process, approach to the teaching of professional genres. *ELT Journal*, 47, 305-316.

Flowerdew, J. (2002). Genre in the classroom: A linguistic approach. In Johns, A. M. (Ed.), *Genre in the classroom: Multiple perspectives*, 91-102.

Flowerdew, J., & Dudley-Evans, T. (2002). Genre analysis of editorial letters to international journal contributors. *Applied Linguistics*, 23(4), 463-489.

Flowerdew, J., & Forest, R. W. (2009). Schematic structure and lexico-grammatical realization in corpus- based genre analysis: The case of research in the PhD literature review. In

M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic Writing: At the Interface of Corpus and Discourse* (pp. 15-36). London: Continnum.

Flowerdew, J., & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes*, 16(1), 27-46.

Flowerdew, J., & Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific purposes*, 29(2), 78-93.

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

Flowerdew, L. (2002). Corpus-based Analyses in EAP, in J. Flowerdew (ed.) *Academic Discourse*, London: Pearson, pp. 95-114.

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Conner & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics*, 11-33. Amsterdam: John Benjamins.

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.

Flowerdew, L. (2016). A genre-inspired and lexico-grammatical approach for helping postgraduate students craft research grant proposals. *English for Specific Purposes*, 42, 1-12.

Fortanet, I. (2004). The use of 'we' in university lectures: reference and function. *English for Specific Purposes*, 23(1), 45-66.

Freedman, A. (1994). "Do as I say?": The relationship between teaching and learning new genres. In A. Freedman & P. Medway (Eds.), *Genre and the New Rhetoric* (pp. 191-210). London: Taylor and Francis.

Freedman, A., & Medway, P. (1994a). *Genre and the new rhetoric*. London: Taylor and Francis.

Freedman, A., & Medway, P. (1994b). Locating genre studies: Antecedents and prospects. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 1-20). London: Taylor & Francis.

Freedman, A., & Medway, P. (1994c). Introduction: New views of genre and their implications for education. In A. Freedman & P. Medway (Eds.), *Learning and teaching genre* (pp. 1-22). Portsmouth, NH: Boynton/Cook.

Geertz, C. (1980). Negara: The theatre state in nineteenth-century Bali. Princeton, NJ: Princeton University Press.

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, 19(2), 115-135.

Goffman, E. (1981). *Forms of talk*. Philadelphia: University of Pennsylvania Press.

Grabowski, Ł., & Juknevičienė, R. (2016). Towards a refined inventory of lexical bundles: an experiment in the Formulex method. *Studies About Languages*, (29), 58-73.

Halleck, G. B., & Connor, U. M. (2006). Rhetorical moves in TESOL conference proposals. *Journal of English for Academic Purposes*, 5(1), 70-86.

Halliday, M. A. K. (1978) *Language as social semiotic: the social interpretation of language and meaning*. London: Edward Arnold.

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

Hammond, K. (2018). "I need it now!" Developing a formulaic frame phrasebank for a specific writing assessment: Student perceptions and recommendations. *Journal of English for Academic Purposes*, 31, 98-105.

Hellekjær, G. O. (2010). Lecture comprehension in English-medium higher education. *Hermes*, 45, 11-34.

Hinkel, E. (2006). *Classroom Talk*. Encyclopedia of Language & Linguistics.

Hirano, E. (2009). Research article introductions in English for specific purposes: Acomparison between Brazilian Portuguese and English. *English for Specific Purposes*, 28(4), 240-250.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.

Holmes, R. (1997). Genre analysis and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, 16(4), 321-337.

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

House, J. (1999). Misunderstanding in intercultural communication: Interactions in English as a lingua franca and the myth of mutual intelligibility. In C, Gnutzmann (Ed.), *Teaching and learning English as a global language*, 7389.

Hsu, W. (2014). The most frequent opaque formulaic sequences in English-medium college textbooks. *System*, 47, 146-161.

Hu, G. (2009). The craze for English-medium education in China: Driving forces and looming consequences. *English Today*, 25(4), 47-54.

Hu, G., & Duan, Y. (2018). Questioning and responding in the classroom: a cross-disciplinary study of the effects of instructional mediums in academic subjects at a Chinese university. *International Journal of Bilingual Education and Bilingualism*, 1-19.

Hu, G., & Lei, J. (2014). English-medium instruction in Chinese higher education: a case study. *Higher Education*, 67(5), 551-567.

Hu, G., & McKay, S. L. (2012). English language education in East Asia: Some recent developments. *Journal of Multilingual and Multicultural Development*, 33, 345-362.

Hu, G., Li, L., & Lei, J. (2014). English-medium instruction at a Chinese University: Rhetoric and reality. *Language Polic*y, 13(1), 21-40.

Hülmbauer, C. (2007). 'You moved, aren't?' — The Relationship between Lexicogrammatical Correctness and Communicative Effectiveness in English as a Lingua Franca. *Vienna English Working Papers*, 16 (2), 3-35.

Hülmbauer, C. (2009). We don't take the right way. We just take the way that we think you will understand-the shifting relationship between correctness and effectiveness in ELF. *English as a lingua franca: Studies and findings*, 323-347.

Humphreys, P. (2017). EMI in Anglophone Nations: Contradiction in Terms or Cause for Consideration?. In Fenton-Smith, B., Humphreys, P., & Walkinshaw, I. (Eds.), *English Medium Instruction in Higher Education in Asia-Pacific* (pp. 93-114). Springer, Cham.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunt, R. (1994). Speech genres, writing genres, school genres, and computer genres. In A. Freedman & P. Medway (Eds.), *Learning and teaching genre* (pp. 243-262). Portsmouth, NH: Boynton/Cook.

Hyland, K. (2002). Genre: Language, context, and literacy. *Annual Review of Applied Linguistics*, 22, 113-135.

Hyland, K. (2004). *Genre and second language writing*. Ann Arbor: The University of Michigan Press.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing, *International Journal of Applied Linguistics,* 18(1): 41-62.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.

Hyland, K. (2012). Bundles in academic discourse. Annual Review of Applied Linguistics, 32. 150-169.

Hyland, K., & Jiang, K. F. (2018). Academic lexical bundles: how are they changing? *International Journal of Corpus Linguistics*, 23(4), 383-407.

Hyon, S. (1996). Genre in three traditions: Implications for ESL. TESOL quarterly, 30(4), 693-722.

Ibrahim, Y. (2001). Doctor and patient questions as a measure of doctor-centredness in UAE hospitals. *English for Specific Purposes*, 20(4), 331-344.

Jarvis, J. and Robinson, M. (1997). Analysing educational discourse: an exploratory study of teacher response and support to pupils' learning, *Applied Linguistics*, 18: 212-228.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied linguistics*, 23(1), 83-103.

Jenkins, J. (2005). Implementing an international approach to English pronunciation: The role of teacher attitudes and identity. *TESOL Quarterly*, 39(3), 535-543.

Jenkins, J. (2011). Accommodating (to) ELF in the international university. *Journal of Pragmatics*, 43(4), 926-936.

Jenkins, J. (2014). *English as a Lingua Franca in the International University. The Politics of Academic English Language Policy*. London: Routledge.

Johns, A. M. (Ed.). (2002). *Genre in the classroom: Multiple perspectives*. NJ: Lawrence Erlbaum Associates.

Kanoksilapatham, B. (2005). Rhetorical move structure of biochemistry research articles. *English for specific purposes*, 24(3), 269-292.

Kanoksilapatham, B. (2011). Civil engineering research article Introductions: Textual structure and linguistic characterization. *The Asian ESP Journal*, 7(2), 55-84.

Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes*, 37, 74-86.

Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16(1), 1-20.

Kashiha, H., & Heng, C. S. (2014). Discourse functions of formulaic sequences in academic speech across two disciplines. *GEMA Online® Journal of Language Studies*, 14(2), 15-27.

Kecskes, I. (2007). Formulaic language in English lingua franca. *Explorations in pragmatics: Linguistic, cognitive and intercultural aspects*, 1, 191-218.

Kirkpatrick, A. (2014). The language (s) of HE: EMI and/or ELF and/or multilingualism?. *The Asian Journal of Applied Linguistics*, 1(1), 4-15.

Klimpfinger, Teresa, 2009. ''She's mixing the two languages together''— forms and functions of code-switching in English as a Lingua Franca. In Mauranen, A., Ranta, E. (Eds.),

English as a Lingua Franca: Studies and Findings. Newcastle: Cambridge Scholars Publishing, pp. 348-371.

Koester, A. (2010). Building small specialised corpora. In A. O'Keeffe & M. McCarthy (Eds.) *The Routledge handbook of corpus linguistics*, 1, 66-79.

Kumaravadivelu, B. (1999). Critical classroom discourse analysis, *TESOL Quarterly*, 33: 453-484.

Kuteeva, M. (2013). Graduate learners' approaches to genre-analysis tasks: Variations across and within four disciplines. *English for Specific Purposes*, *32*(2), 84-96.

Kuteeva, M., & Negretti, R. (2016). Graduate students' genre knowledge and perceived disciplinary practices: Creating a research space across disciplines. *English for specific purposes*, *41*, 36-49.

Kwan, B. S. C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, 25, 30-55.

Lantolf, J.P. (2000). *Sociocultural Theory and Second Language Learning*, Oxford: Oxford University Press.

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, 39, 45-61.

Lebauer, R. S. (1984). Using lecture transcripts in EAP lecture comprehension courses. *TESOL Quarterly*, 18(l), 41-53.

Lee, J. J. (2009). Size matters: an exploratory comparison of small-and large-class university lecture introductions. *English for Specific Purposes*, 28(1), 42-57.

Lee, J. J. (2011). *A genre analysis of second language classroom discourse: Exploring the rhetorical, linguistic, and contextual dimensions of language lessons*. Unpublished Doctoral Dissertation, Georgia State University.

Lee, J. J. (2016). "There's intentionality behind it…": A genre analysis of EAP classroom lessons. *Journal of English for Academic Purposes*, 23, 99-112.

Lee, J. J., & Subtirelu, N. C. (2015). Metadiscourse in the classroom: A comparative analysis of EAP lessons and university lectures. *English for Specific Purposes*, 37, 52-62.

Levinson, S. (1992). Activity types and language. In Drew, P., Heritage, J. (Eds.), *Talk at Work: Interaction in Institutional Settings*. Cambridge University Press, Cambridge, pp. 66-100.

Li, H. (2010). Contrastive genre analysis of English and Chinese social science research article introductions. *Journal of Chongqing University of Technology (Social Science)*, 24(11), 110-113.

Liang, M., Li, W. & Xu J. (2010). *Using corpora: A practical coursebook*. Beijing: Foreign Language Teaching and Research Press.

Lim, J. M. H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes*, 25(3), 282-309.

Lim, J. M. H. (2010). Commenting on research results in applied linguistics and education: A comparative genre-based investigation. *Journal of English for academic Purposes*, *9*(4), 280-294.

Lim, J.M. H. (2011). 'Paving the way for research findings': Writers' rhetorical choices in education and applied linguistics. *Discourse Studies*, 13(6), 725-749.

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes*, *31*(3), 150-160.

Literacy and Education Research Network. (1990). *A genre-based approach to teaching writing, years 3-6: Book 1. Introduction. Annandale*, Australia: Common Ground.

Loi, C. K. (2010). Research article introductions in Chinese and English: A comparative genre-based study. *Journal of English for Academic Purposes*, *9*(4), 267-279.

Loi, C. K., & Evans, M. S. (2010). Cultural differences in the organization of research article introductions from the field of educational psychology: English and Chinese. *Journal of Pragmatics*, 42(10), 2814-2825.

Lorés-Sanz, R. (2016). ELF in the making? simplification and hybridity in abstract writing. *Journal of English as a Lingua Franca*, *5*(1), 53-81.

Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, 39, 21-36.

Lu, X., Yoon, J., & Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes*, 36, 76-85.

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, *51*(1), 36-76.

Malinowski, B. (1935). *Coral gardens and their magic: a study of methods of tilling the soil and agricultural rites in the Trobriand Islan*ds. Language of Trobriand Islands.

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for Specific Purposes*, 19(1), 63-86.

Martin, J. R., Christie, F., & Rothery, J. (1987). Social processes in education: A reply to Sawyer and Watson (and others). In I. Reid (Ed.), *The place of genre in learning: Current debates* (pp. 46-57). Geelong, Australia: Deakin University Press.

Martín, P. M. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for specific purposes*, *22*(1), 25-43.

Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT journal*, 67(2), 184-198.

Martinez, R. (2018). "Specially in the last years…": Evidence of ELF and non-native English forms in international journals. *Journal of English for Academic Purposes*, 33, 40-52.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. Applied linguistics, 33(3), 299-320.

Mauranen, A. (2003). The corpus of English as lingua franca in academic settings. *TESOL quarterly*, 37(3), 513-527.

Mauranen, A. (2009). Chunking in ELF: Expressions for managing interaction. *Intercultural Pragmatics*, 6(2), 217-233.

Mauranen, A. (2012). *Exploring ELF: Academic English Shaped by Non-native Speakers*. Cambridge: Cambridge University Press.

Mauranen, A. (2014). Lingua Franca discourse in academic contexts: Shaped by complexity. In Flowerdew, J. (Ed.) *Discourse in Context: Contemporary Applied Linguistics*. Vol. 3. London: Bloomsbury, pp. 225-245.

McEnery, A. M., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. New York: Taylor & Francis.

McKenna, B. (1997). How engineers write: An empirical study of engineering report writing. *Applied Linguistics*, 18(2), 189-211.

Meierkord, C. (2013). Interpreting successful lingua-franca interaction. An analysis of non-native-/non-native small talk conversation in English. *Linguistik online*, 5(1).

Metsä-Ketelä, M. (2006). "Words are more or less superfluous": the Case of more or less in Academic Lingua Franca English. *Nordic Journal of English Studies*, 5(2), 117-43.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, *32*, 111-129.

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

Miller, C. R. (1994). Rhetorical community: The cultural basis of genre. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 67-78). London: Taylor & Francis.

Ministry of Education. (2001). *Guanyu jiaqiang gaodeng xuexiao benke jiaoxue gongzuo tigao jiaoxue zhiliang de ruogan yijian [Guidelines for improving the quality of undergraduate teaching]*. Beijing: Author.

Ministry of Education. (2004). *Putong gaodeng xuexiao benke jiaoxue gongzuo shuiping pinggu fangan [Assessment standards for undergraduate education in higher institutions]*. Beijing: Author.

Ministry of Education. (2005). *Guanyu jinyibu jiaqiang gaodeng xuexiao benke jiaoxue gongzuo de ruogan yijian [Guidelines for further improving undergraduate education]*. Beijing: Author.

Ministry of Education. (2007). *Guanyu jinyibu shenhua benke jiaoxue gaige quanmian tigao jiaoxue zhiliang de ruogan yijian [Guidelines for deepening the reform of undergraduate education and raising educational quality]*. Beijing: Author.

Ministry of Education. (2017). Quanguo gaodeng xuexiao mingdan *[National list of colleges and universities]*. http://www.moe.gov.cn/srcsite/A03/moe_634/201706/t20170614_306900.html Accessed October 17, 2018

Mitchell, N. (2016). *Universities around the world offer tuition in English*. www.bbc.co.uk . Accessed July 25, 2016.

Mizumoto, A., Hamatani, S., & Imao, Y. (2017). Applying the bundle-move connection approach to the development of an online writing support tool for research articles. *Language Learning*, 67(4), 885-921.

Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, 50, 40-63.

Moritz, M. E. W., Meurer, J. L., & Dellagnelo, A. K. (2008). Conclusions as components of research articles across Portuguese as a native language, English as a native language

and English as a foreign language: A contrastive genre study. T*he ESPecialist*, 29, 233-253.

Morley, J. (2014). *Academic phrasebank*. Manchester: University of Manchester.

Morton, J. (2009). Genre and disciplinary competence: A case study of contextualisation in an academic speech genre. *English for Specific Purposes*, *28*(4), 217-229.

Namba, K. (2008). Formulaic language in bilingual children's code-switching. Unpublished PhD thesis, Center for language and communication Research, Cardiff University.

Nattinger, J. (1986). Lexical phrases, functions and vocabulary acquisition. *The ORTESOL Journal,* 7, 1-14.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford University Press.

Neely, E., & Cortes, V. (2009). A little bit about: analyzing and teaching lexical bundles in academic lectures. Language Value, 1, 17-38.

Nekrasova, T. M. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language learning*, 59(3), 647-686.

Nesi, H. & Basturkmen, H. (2009). Lexical bundles and discourse signaling in academic lectures. In J. Flowerdew & M. Mahlberg (eds.), *Lexical Cohesion and Corpus Linguistics*, 23-43. Amsterdam and Philadelphia: John Benjamins.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), 283-304.

Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd Edition). Los Angeles: Sage

Publications.

Nwogu, K. N. (1991). Structure of science popularizations: A genre-analysis approach to the

schema of popularized medical texts. *English for Specific Purposes*, 10(2), 111-123.

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for specific

purposes*, 16(2), 119-138.

O'Keeffe, A., & McCarthy, M. (Eds.). (2010). *The Routledge handbook of corpus linguistics*.

Routledge.

Okamura, A. (2009). Use of personal pronouns in two types of monologic academic speech.

The Economic Journal of Takasaki City University of Economics, 52(1), 17-26.

Olsen, L. A., & Huckin, T. H. (1990). Point-driven understanding in engineering lecture

comprehension. *English for Specific Purposes*, 9(1), 33-47.

Ozturk, I. (2007). The textual organisation of research article introductions in applied

linguistics: Variability within a single discipline. *English for Specific Purposes*,

26(1), 25-38.

Paltridge, B. (1994). Genre analysis and the identification of textual boundaries. *Applied

Linguistics*, 15(3), 288-299.

Paltridge, B. (2001). *Genre and the language learning classroom*. Ann Arbor: University of

Michigan Press.

Paltridge, B. (2012). *Discourse analysis: An introduction*. Bloomsbury Publishing.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.

Parkinson, J. (2011). The discussion section as argument: The language used to prove knowledge claims. *English for Specific Purposes*, 30, 164-175.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, 191-226.

Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System*, 30(4), 479-497.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94.

Peters, A. M. (1983). *Units of language acquisition*. Cambridge: Cambridge University press.

Peters, E., & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for academic purposes*, 20, 28-39.

Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, 10, 231-250.

Piller, I., & Cho, J. (2013). Neoliberalism as language policy. *Language in Society*, 42, 23-44.

Posteguillo, S. (1999). The semantic structure of computer science research articles. *English for Specific Purposes*, 18(2), 139-160.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42, 220-231.

Ranta, E. (2006). The 'Attractive' Progressive-Why use the-ing Form in English as a Lingua Franca?. *Nordic Journal of English Studies*, 5(2), 95-116.

Read, J. (2015). *Assessing English proficiency for university study*. London, UK: Palgrave Macmillan.

Richards, J. C., & Lockhart, C. (1996). Interaction in the second language classroom. *Readings in Methodology*, 66-70.

Rounds, P. L. (1987). Multifunctional personal pronoun use in an educational setting. *English for Specific Purposes*, 6, 13-29.

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

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Philadelphia, PA: John Benjamins.

Schmid, H. (2000). *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin: Walter de Gruyter.

Schmitt, N. & N. Carter. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic Sequences Acquisition, processing and use*. Amsterdam: John Benjamins, pp. 1-22.

Schnur, E. (2014). Phraseological signaling of discourse organization in academic lectures: A comparison of lexical bundles in authentic lectures and EAP listening materials. *Yearbook of Phraseology*, 5(1), 95-122.

Schryer, C. F. (1993). Records as genre. *Written Communication*, 10, 200-234.

Seedhouse, P. (2004). *The Interactional Architecture of the Second Language Classroom: A Conversational Analysis Perspective*, Oxford: Blackwell.

Seedhouse, P. (2010). The interactional architecture of the language classroom. *Bellaterra Journal of Teaching & Learning Language & Literature*, 1(1), 1-13.

Seidlhofer, B. (2004). 10. Research perspectives on teaching English as a lingua franca. *Annual review of applied linguistics*, 24, 209-239.

Seidlhofer, B. (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Shi, H-M (2014). *Moves and lexical bundles: A contrastive study of English agricultural science research articles between Chinese journals and internationally published journals*. Unpublished Doctoral Dissertation, Suranaree University of Technology.

Simpson, R. C., Lee, D. Y. W., & Leicher, S. (2007). *MICASE Manual: The Michigan Corpus of Academic Spoken English [online] Version 1.1*. Michigan: The English Language Institute, University of Michigan. Accessed May 19, 2018.

Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *Tesol Quarterly*, *37*(3), 419-441.

Simpson-Vlach, R. C., & Leicher, S. (2006). The MICASE handbook: A resource for users of the Michigan corpus of academic spoken English. Ann Arbor: University of Michigan Press.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. M. (2004). The search for units of meaning. In *Trust the text: Language, corpus and discourse* (pp. 24-48). London: Routledge.

Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.

Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon*, 8(2), 245-268.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251-272.

Spolsky, B. (2004). *Language policy*. Cambridge, UK: Cambridge University Press.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214-225.

Stoller, F. L., & Robinson, M. S. (2013). Chemistry journal articles: An interdisciplinary approach to move analysis with pedagogical aims. *English for Specific Purposes*, 32(1), 45-57.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. *Functions of language*, 2, 1-33.

Suárez, L., & *A*. Moreno *(*2008). The rhetorical structure of academic book reviews of literature: An English-Spanish cross-linguistic approach. In U. Connor, E. Nagelhout & W. Rozycki (Eds.). *Contrastive Rhetoric: Reaching to intercultural rhetoric* (pp. 147-168). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Suviniitty, J. (2010). Lecturers' questions and student perception of lecture comprehension. *Helsinki English Studies*, 6(1), 44-57.

Swales, J. M. (1981). *Aspects of article introductions* (Aston ESP Research Report 1). Birmingham, England: University of Aston in Birmingham, Language Studies Unit.

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

Swales, J. M. (1998). *Other floors, other voices: A textography of a small university building*. Mahwah, NJ: Lawrence Erlbaum.

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

Swales, J., & Feak, C. (1994). *Academic writing for graduate students: Essential tasks and skills*. Ann Arbor: University of Michigan Press.

Swales, J., & Feak, C. (2000). English in today's research world: A writing guide. Ann Arbor: University of Michigan Press.

Taguchi, N. (2014). English-medium education in the global society. *International Review of Applied Linguistics in Language Teaching*, 52(2), 89-98.

Tange, H. (2012). Organising language at the international university: Three principles of linguistic organisation. *Journal of Multilingual and Multicultural Development*, 33, 287-300.

Taylor, J. R. (2006). Polysemy and the lexicon. *Applications of Cognitive Linguistics*, 1, 51-80.

Tessuto, G. (2015). Generic structure and rhetorical moves in English-language empirical law research articles: Sites of interdisciplinary and interdiscursive cross-over. *English for Specific Purposes*, *37*, 13-26.

Thompson, S. E. (1994). Frameworks and contexts: A genre-based approach to analyzing lecture introductions. *English for Specific Purposes*, 13(2), 71-86.

Thompson, S. E. (2003). Text-structuring metadiscourse, intonation and the signalling of organisation in academic lectures. *Journal of English for academic purposes*, 2(1), 5-20.

Thompson, Susan (1998). Why ask questions in monologue? Language choices at work in scientific and linguistic talk. In Susan Hunston (Ed.) *Language at work*, Clevedon: BAAL/Multilingual Matters, 137-150.

Thornborrow, J. (2001). Questions, control and the organization of talk in calls to a radio phone-in. *Discourse Studies*, 3(1), 119-143.

Tong, F., & Shi, Q. (2012). Chinese-English bilingual education in China: A case study of college science majors. *International Journal of Bilingual Education and Bilingualism*, 15(2), 165-182.

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131-149). London: Longman, Pearson Education.

Uhrig, K. (2012). Business and legal case genre networks: Two case studies. *English for Specific Purposes*, *31*(2), 127-136.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: an eye-movement study into the processing of formulaic sequences' in N. Schmitt (ed.): *Formulaic Sequences*. John Benjamins. 153-73.

Unterberger, B. (2012). English-medium programmes at Austrian business faculties: A status quo survey on national trends and a case study on programme design and delivery. *AILA Review*, 25, 80-100.

Upton, T. A. (2002). Understanding direct mail letters as a genre. *International Journal of Corpus Linguistics*, *7*(1), 65-85.

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.

van Lier, L. (1988). *The classroom and the language learner: Ethnography and second-language classroom research*. London: Longman.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge: Harvard University Press.

Walkinshaw I., Fenton-Smith B., Humphreys P. (2017) EMI Issues and Challenges in Asia-Pacific Higher Education: An Introduction. In: Fenton-Smith B., Humphreys P., Walkinshaw I. (Eds), *English Medium Instruction in Higher Education in Asia-Pacific. Multilingual Education*, vol 21 (pp. 1-18). Cham: Springer.

Walsh, S. (2011). *Exploring classroom discourse: Language in action*. London: Taylor & Francis.

Wang, Y. (2017). Lexical bundles in spoken academic ELF: Genre and disciplinary variation. *International Journal of Corpus Linguistics*, 22(2), 187-211.

Wang, Y. (2018a). As Hill seems to suggest: Variability in formulaic sequences with interpersonal functions in L1 novice and expert academic writing. *Journal of English for Academic Purposes*, 33, 12-23.

Wang, Y. (2018b). Formulaic sequences signalling discourse organisation in ELF university lectures: A disciplinary perspective. *Journal of English as a Lingua Franca*, 7(2), 355-376.

Wang, Y. (2019). A functional analysis of text-oriented formulaic expressions in written academic discourse: Multiword sequences vs. single words. *English for Specific Purposes*, *54*, 50-61.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied linguistics*, 16(2), 180-205.

Weissberg, R., & Buker, S. (1990). *Writing up research: Experimental report writing for students of English*. Englewood Cliffs, NJ: Prentice-Hall Regents.

Wennerstrom, A. (2003). *Discourse analysis in the classroom. Volume 2. Genres of writing*. Ann Arbor: University of Michigan Press.

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

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

Wilkinson, R. (2013). English-medium instruction at a Dutch university: Challenges and pitfalls. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 3-24). Bristol, UK: Multilingual Matters.

Wolff, D. 2007. CLIL: Bridging the gap between school and working life. In D. Marsh & D. Wolff (Eds), *Diverse Contexts -Converging Goals: CLIL in Europe*, 15-25. Frankfurt: Peter Lang.

Wong Fillmore, L. (1985). When does teacher talk work as input? In S. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 17e50). Rowley, MA: Newbury House.

Wood, D. (2002). Formulaic language acquisition and production: Implications for teaching. *TESL Canada Journal*, 01-15.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*, 21(4), 463-489.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, *20*(1), 1-28.

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, 54, 1-14.

Wu, P., S. Wang, X. Jiang, D. Zeng, Y. Guan, and X. Li. (2010). *Gaodeng Xuexiao shuangyu jiaoxue de xianzhuang yanjiu he shijian taolun* [An Exploratory Study of English-medium Instruction Practices in Chinese Higher Education Institutions]. Beijing: Higher Education Press.

Xu, J., & Jia, Y. (2011). BFSU qualitative coder (Version 1.1) [Computer Software]. Bejing: National Research Center for Foreign Language Education, Beijing Foreign Studies University.

Yakhontova, T. (2002). "Selling" or "telling"? The issue of cultural variation in research genres. In J. Flowerdew (Ed.), *Academic discourse* (pp. 216-232). London: Longman.

Yang, R-Y. & D. Allison (2003). Research articles in applied linguistics: moving from results to conclusions. *English for Specific Purposes*, 22(4), 365-385.

Yeo, J. Y., & Ting, S. H. (2014). Personal pronouns for student engagement in arts and science lecture introductions. *English for Specific Purposes*, 34, 26-37.

Young, L. (1994). University lectures - macro-structure and micro-features. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 159-175). Cambridge: Cambridge University Press.

Zhao, J., and L. Q. Dixon. (2017). Introduction. In Zhao, J., & Dixon, L. Q. (Eds.), *English-medium Instruction in Chinese Universities: Perspectives, Discourse and Evaluation*, (pp.1-19). New York: Routledge.

Zhu, Y. (1997). An analysis of structural moves in Chinese sales letters. *Text*, 17, 543-566.

Zhu, Y., & Yu, L. (2010). Woguo hanying shuangyu jiaoyu yanjiu xianzhuang [The state-of-the-art of English-Chinese bilingual education in China]. *Foreign Languages and Their Teaching*, 255, 31-36.

APPENDICES
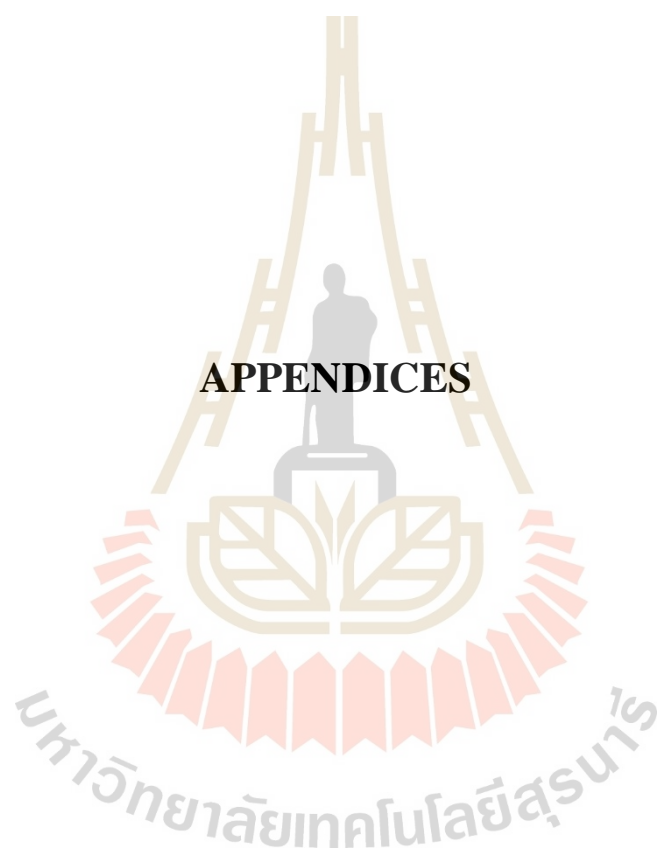
APPENDICES

APPENDIX A

General information of Corpus of English Lecturers (CEL)

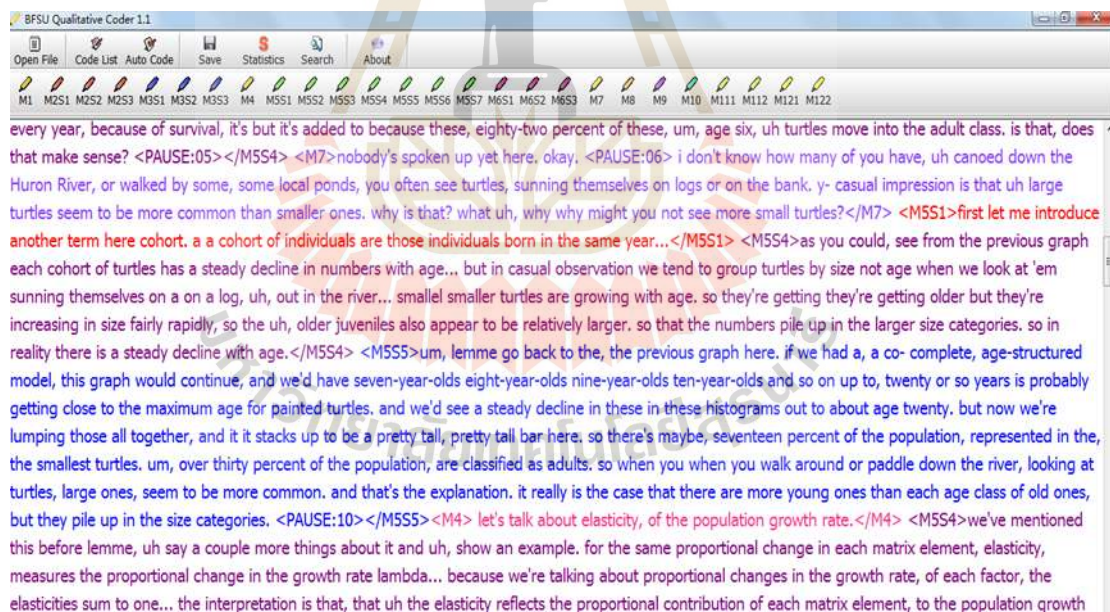| No.	Title	Source	Academic Division	Duration (minutes)	Word count	Number of students
CEL01	Fundamentals of radiation chemistry	BASE_pslct005	Chemistry	49	7496	10
CEL02	Rigid body dynamics	BASE_pslct018	Engineering	49	6751	20
CEL03	Polymers	BASE_pslct019	Engineering	51	7048	16
CEL04	Precipitation	BASE_pslct028	Meteorology	45	5924	40
CEL05	Methods and developments in plant breeding	BASE_lslct001	Agricultural Botany	55	9616	5
CEL06	HIV and AIDS	BASE_lslct008	Biological science	49	6276	<20
CEL07	The Science of Transplantation	BASE_lslct011	Biological Sciences	99	15713	40
CEL08	Biology and Ecology of Fishes Lecture	MICASE_LES175SU025	Biological and Health Sciences	70	9708	10
CEL09	Biology of Birds Lecture	MICASE_LES175SU028	Biological and Health Sciences	84	12212	17
CEL10	Graduate Industrial Operations Engineering Lecture	MICASE_LES330JG052	Physical Sciences and Engineering	81	10939	15
CEL11	Graduate Population Ecology Lecture	MICASE_LES425JG077	Biological and Health Sciences	44	5312	18
CEL12	Intro Programming Lecture	MICASE_LES235SU099	Physical Sciences and Engineering	50	8057	17

APPENDIX B

General information of Corpus of Chinese Lecturers (CCL)

No.	Title	Source	Academic Division	Duration (minutes)	Word count	Number of students
CCL01	Geologic Time: Earth's Evolution	Xi'an Shiyou University	School of Earth Sciences and Engineering	45	3785	10
CCL02	Igneous Rocks	Xi'an Shiyou University	School of Earth Sciences and Engineering	47	4606	5
CCL03	An Outline of Earth Science	Xi'an Shiyou University	School of Earth Sciences and Engineering	50	4109	16
CCL04	Introduction to Physical Geology 一 Sediment Transport and Deposition	Xi'an Shiyou University	School of Earth Sciences and Engineering	48	4735	9
CCL05	An Outline of Earth Science	Xi'an Shiyou University	School of Earth Sciences and Engineering	45	4524	10
CCL06	Gas Well Unloading Technologies	Xi'an Shiyou University	College of Petroleum Engineering	53	6807	11
CCL07	Recent Advances in Oil and Gas Production Engineering- Introduction	Xi'an Shiyou University	College of Petroleum Engineering	46	4911	14
CCL08	Advanced Hydraulic Fracturing Technologies	Xi'an Shiyou University	College of Petroleum Engineering	51	7073	14
CCL09	Reservoir Engineering	Xi'an Shiyou University	College of Petroleum Engineering	41	4552	4
CCL10	Reservoir Engineering overview	Xi'an Shiyou University	College of Petroleum Engineering	54	3917	4
CCL11	The Derivative	Xi'an Shiyou University	School of Science	46	4473	12
CCL12	Infinite Sequences and Its Limits	Xi'an Shiyou University	School of Science	46	3064	14

APPENDIX C

SAMPLE OF MOVE-CODED TEXT SEGMENT AND THE OPERATING WINDOW OF BFSU QUALITATIVE CODER 1.1

<M3S1>today we're gonna be talking about, stage-structured and size-structured models. again, uh matrix models, following up on the, the, topics we've uh, discussed earlier.</M3S1> <M2S1>mostly we've been talking about, age-structured models or it's Leslie matrix models.</M2S1> <M3S1>now we're going to um, talk about some variations on that theme. <PAUSE:06></M3S1> <M4>first we're gonna talk about, a relatively simple, modification of an age-based or Leslie matrix model.</M4>

APPENDIX D

SAMPLE OF MOVE STATISTICS

Item(s): 26 Hit(s): 18					
No	**Name**	**Tag Set**	**Freq.**	**Count**	**Tagged Text(s)**
1	**M1**	<M1>...</M1>	0	0	
2	**M2S1**	<M2S1>...</M2S1>	1	1	mostly we've been talking about, age-structured models or it's Leslie matrix models.
3	**M2S2**	<M2S2>...</M2S2>	0	0	
4	**M2S3**	<M2S3>...</M2S3>	0	0	
5	**M3S1**	<M3S1>...</M3S1>	2	2	today we're gonna be talking about, stage-structured and size-structured models. again, uh matrix models, following up on the the, the, topics we've uh, discussed earlier. now we're going to um, talk about some variations on that theme.
6	**M3S2**	<M3S2>...</M3S2>	1	1	for long-lived species, an aged-based Leslie matrix model can be very large. for example uh, some of the sea turtles, can live to over fifty years. so if you have uh, an um, Leslie matrix that has one entry for each, year class that can be a fifty by fifty matrix. which is not only, w- uh, a big matrix, and uh, m- moderately cumbersome to to work with. it can be hard to uh, supply the data for that. um, if you wanna estimate the transitions for, specific age classes, you'll need some replication. so you need to be able to age the animals well and get sufficient, uh data on each age class to accurately estimate, the transition. often that's quite hard and so simplifications are made to, to average use average survival rates for the older age classes. and sometimes for younger age classes as well. um, also for long-lived species, the oldest age can be hard to measure accurately... if the population of is of relatively modest size, and especially if it's if it's difficult to measure ages, it can be h- it can be hard to tell, what the actual oldest age is, uh is likely to be just because there's, can be a lotta chance variations in the survival especially of the older groups and there's gonna be small numbers. so it can take a while before you g- you um, very accurately identify, what the oldest age is and therefore what the, the size of your life table and your, your matrix model should be...
7	**M3S3**	<M3S3>...</M3S3>	0	0	
8	**M4**	<M4>...</M4>	7	7	the last column shows the reproductive value for an individual by stage. we can do an eigenvalue analysis with the, stage-based models just like we did with the age-based models... we can still use our, uh our m- matrix analysis techniques. and calculate the stable age distribution, of painted turtles, let's uh lemme get back to, first we're gonna talk about, a relatively simple, modification of an age-based or Leslie matrix model. i hope you can see that the picture okay. one of the first stage-structured models for plants was developed for teasel. Dipsacus sylvestris and it's in the teasel family. let's talk about elasticity, of the population growth rate.

APPENDIX E
GUIDING QUESTIONS FOR INTERVIEWS

1. Do you follow similar organizational patterns when you give EMI lectures and Chinese lectures?

2. Do you spend time on classroom management at the beginning of your EMI lectures?

3. Do you make explicit proclamation of lecture scope to students?

4. Do you prefer elaborating topics in depth or talk in generalities?

5. How do you introduce technical terms and concepts to students?

6. How do you present background knowledge to students?

7. Do you prefer to quote authorities while lecturing? Why?

8. Do you present both the main points as well as the caveats to students? Why?

9. Do you make comments on lecture contents? Why?

10. Do you often clarify yourself while lecturing? Why?

11. Are you concerned with creating connections between knowledge points? How do you do that?

12. Do you recommend more reading materials while lecturing?

13. Do you check students understanding while lecturing? If so, how do you do that?

14. Do you prefer to end the lecture with a summary of major contents?

15. Do you think Chinese EMI lecturers possess a more limited repertoire of FSs than native English lecturers?

16. Do you realize some of the FSs you use are non-standard (e.g., *because…so*, *three kind of* and *is depends on*)? Do you mind using such non-standard FSs?

CURRICULUM VITAE

Mr. Lin Deng was born in Jiangsu Province, China in September, 1980. He received his B.A. in English from Xi'an Shiyou University in 2003. Since then, he has been teaching English in Xi'an Shiyou University. In 2011, he obtained his M.A in Linguistics and Applied Linguistics from Xi'an International Studies University. From 2015 to 2020, he studied for his Ph.D. in English language studies in School of Foreign Languages, Institute of Social Science Technology, Suranaree University of Technology, Thailand. He is currently an associate professor in the School of Foreign Languages, Xi'an Shiyou University. His research interests include genre analysis and corpus linguistics.