MOVES IN DISCUSSIONS: A CORPUS-BASED GENRE ANALYSIS OF THE

DISCUSSION SECTIONS IN APPLIED LINGUISTICS RESEARCH ARTICLES

WRITTEN IN ENGLISH

Dissertation Submitted to the Institute of Educational Sciences

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

In

English Language Teaching

Evrim Eveyik-Aydın

YEDITEPE UNIVERSITY

2015

YEDİTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ TESLİM ve ONAY TUTANAĞI

KONU: Moves in Discussions: A Corpus-based genre analysis of the discussion sections in Applied Linguistics Research Articles written in English

ONAY:

Yrd. Doç. Dr. Erkan Karabacak
(Danışman)   (İmza)

Prof. Dr. Ayşe S. Akyel
(Danışman)   (İmza)

Hossein Farhady
(Üye)   (İmza)

Doç. Dr. Çiler Hatipoğlu
(Üye)   (İmza)

Deniz Tat
(Üye)   (İmza)

TESLİM EDEN   : Evrim Eveyik-Aydın
TEZ SAVUNMA TARİHİ   : 27. 03. 2015
TEZ ONAY TARİHİ   : 9. 04. 2015

ii

CURRICULUM VITAE

Evrim Eveyik-Aydın

BA: Bogazfici University, English Language Teaching, 1995

MA: Bogazici University, English Language Education, 1999

PhD: Yeditepe University, English Language Education, 2015

Email: evrimaydin@yeditepe.edu.tr

evrimev@gmail.com

Teaching Experience	**Teacher Trainer**

Teacher Trainer

Yeditepe University, The School of Education, Department of Foreign Language Education   Fall 2010-present
Full-time instructor

Yeditepe University, The School of Education, Department of Foreign Language Education   Summer 2006- Spring 2007
Part-time instructor

Bogazici University, The School of Education, Department of Foreign Language Education   Spring 2006
Part-time instructor

English Teacher
Kalamis High School, Teaching English as a Foreign Language
                    Fall 1995-Spring 1996

ESP Instructor
General English/English for Specific Purposes (ESP) to adults at intermediate level   Spring/Summer 1996

Graduate Assistantships   **E-ELT Project Assistant/**   Fall 2007-2010
Research Assistant
Internet-based MA program in English Language Teaching for in-service teachers
Yeditepe University, Istanbul, Turkey

Administrative Director Assistant   Spring 2002-Fall 2003
Electrical and Computer Engineering Department, WEMPEC, Univesity of Wisconsin-Madison, WI, USA.
Assisted Administrative Director with the organization of annual conferences, meetings, and publications.

Teaching Assistant
University of Wisconsin-Madison, USA, Department of English, Program in English as a Second Language   Fall 2000
(Courses taught: ESL Reading)

Research Assistant   Fall 1996-Fall 1999
The School of Education, Department of Foreign Language Education, Boğaziçi University, Istanbul, Turkey

Projects Contributed   Internet-based MA program on English Language Teaching offered by English Language Teaching Program at Yeditepe University (pilot study)

Session chair at the 20th Annual Second Language Research Forum at the University of Wisconsin-Madison, Sep.7-10, 2000

Assisted with the implementation of senior students' practice teaching in high schools in Partial Fulfillment of the Requirement for the Bogazici University Certificate in TEFL (1996-1999).

Improvement of Evaluation Criteria for Senior Students' Teaching Performance in the Classroom during their Practice Teaching in Partial Fulfillment of the Requirement for the Bogazici University Certificate in TEFL (with Eileen S.Yalcın/ 1996)

Conference Presentations   Eveyik-Aydin, E. (2014, June). *Teachers' classroom management beliefs and practices: The case of a reflective and non-reflective teacher*. Paper presented at the Conference of Teacher Researchers in Action, İzmir, Türkiye.

Eveyik-Aydin & Kesli, Y. (2009, May). *An oral communication syllabus designed for freshman year ELT students.* Paper presented at I. Uluslararası Türkiye Eğitim Araştırmaları Kongresi, Çanakkale, Türkiye.

Eveyik-Aydin, E., Kurt, G. & Mede, E. (2009, February). *Exploring the relationship between teacher beliefs and styles on classroom management in relation to actual teaching practices: A case study.* Paper presented at the World Conference on Educational Sciences, Nicosia, North Cyprus.

Camlibel, Z., & Eveyik-Aydin, E. (2004, May). *Students' views on the*

medium of instruction in EFL countries. Paper presented at the annual meeting of the American Association for Applied Linguistics, Portland, OR.

Gokcora, D., & Eveyik-Aydin, E. (2003, November). *Arabic instructors' perceptions of teaching Arabic communicatively.* Paper presented at the 37th annual meeting and exposition of American Council on the Teaching of Languages (ACTFL), Philadelphia, PA.

Gokcora, D., & Eveyik-Aydin, E. (2003, May). *Perceptions of language instructors on communicative language teaching.* Paper presented at the 3rd International Conference on Language Teacher Education (ICLTE), Minneapolis, MN.

Gokcora, D., & Eveyik-Aydin, E. (2003, May). *Arabic instructors' attitudes towards communicative language teaching*. Paper presented at the annual conference of the National Council of Organizations of Less Commonly Taught Languages (NCOLCTL), Los Angeles, CA.

Eveyik-Aydin, E. (2003, March). *EFL teachers' voice on communicative language teaching.*Paper presented at the annual convention of Teaching English to the Speakers of Other Languages (TESOL), Baltimore, Maryland.

Eveyik-Aydin, E. (2002, April). *EFL teachers' attitudes towards communicative language teaching.* Paper presented at the annual convention of Teaching English to the Speakers of Other Languages (TESOL), 2002, Salt Lake City, Utah.

Eveyik, E. (1999, April). *Development of an attitude scale: Validity and reliability studies.*Paper presented at the Applied Linguistics Colloquium, University of Wisconsin-Madison.

Publications, Theses
Eveyik-Aydin, E. (2015). Moves in discussions: A corpus-based genre analysis of the discussion sections in Applied Linguistics research articles written in English. PhD Dissertation, Yeditepe University, Istanbul, Turkey.

Gokcora, D. & Eveyik-Aydin, E. (2011). *Arabic instructors' attitudes on communicative language teaching: Perspectives from an intensive language program.* LAP Lambert Academic Publshing

Eveyik-Aydin, E., Kurt, G. & Mede, E. (2009). Exploring the relationship between teacher beliefs and styles on classroom management in relation to actual teaching practices: A case study. *Procedia- Social and Behavioral Sciences,1*(1), 618-622.

Gokcora, D., & Eveyik-Aydin, E. (2004). Arabic instructors' attitudes on communicative language teaching (CLT). *Journal of National Council of Less Commonly Taught Languages, 1*, 138-176.

Eveyik-Aydin, E. (2003). EFL teachers' voice on communicative language teaching. *CD-ROM of the annual convention of Teaching English to the Speakers of Other Languages (TESOL).* (ERIC Document Reproduction Service No. ED 476748)

Eveyik, E. (1999). *Development of an attitude scale to investigate Turkish EFL teachers' attitudes towards some aspects of communicative language teaching.* Unpublished Master's Thesis, Bogazici University, Istanbul, Turkey.

Professional Development

Post-MA studies at the Program in English as a Second Language, University of Wisconsin-Madison, (2000-2001)

Took online courses offered by the Program in Teacher Leadership, University of Illinois at Springfield (Summer-Fall 2004)

Attended CARLA (Center for Advanced Research on Language Acquisition) Summer Institute for Language Teachers on "Using Technology in the Second Language Classroom" at the University of Minnesota, August 11-15, 2003, Minneapolis, MN, USA.

Attended Software Training Courses offered by the Division of Information Technology (DoIT) at the University of Wisconsin-Madison (Spring 2001-Fall 2002).

Languages

Turkish (native)
English (advance)
French (intermediate reading skills)

Research Intererests

Corpus Linguistics, genre analysis, teacher education, materials development, second/foreign language teaching methodology

Interests and Activities

Oil painting on canvas, classical music, traveling, scrapbooking, swimming, and reading on art, civilizations, education, history and psychology.
Assistant teacher at the Theatre Club in Kalamis High School, 1995-1996
Committee member of Education and Research Club (EREC), Bogazici University, 1991-1995
Member of News broadcasting Committee at Radio Bogazici, 1994-1995

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my heartfelt gratitude to those who endlessly supported me to complete my doctoral studies successfully.

First of all, I thank my mentor and advisor Prof. Ayşe S. Akyel who genuinely believed in me when setting out on this path and encouraged me constantly to achieve more than I though possible all through my academic studies. I am truly grateful for her years of guidance, moral support and reassuarance, for her exceptional professionalism that enlightened me not only as a graduate student but also as a teacher, and for her invaluable insights that led me into success during this research. Having been a great source of inspiration, and a caring master to look up to, she has always been more than a supervisor for me.

This study would not have been possible without the extraordinary help, remarkable guidance and inspiratinal suggestions of my advisor Assist. Prof. Erkan Karabacak who so generously spent hours with me during the most labour-intense analyses of the study. Throughout this research, I learned tremendously from his expertise on corpus linguistics, and I truly felt I would have been lost in the sea of data without him. Therefore, I present him the most sincere thanks from the bottom of my heart.

I am also indebted to my distinguished committee members Prof. Hossein Farhady, Assoc. Prof. Çiler Hatipoğlu, and Assist. Prof. Deniz Tat for their invaluable comments that improved my study.

I owe much gratitude to my dearest friends and colleagues Yelkin Diker Coşkun, Gonca Kızılkaya Cumaoğlu, Çağla Garipağaoğlu, Görsev and Volkan İnceçay, Aslı Bağışlayıcı, Merve Selçuk and Kenan Dikilitaş for their hearty and invaluable friendship, ever-present encouragement, and their confidence in me. I also express my sincere gratitude

To my parents, husband, and beloved sons Erke and Efe

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

MOVES IN DISCUSSIONS: A CORPUS-BASED GENRE ANALYSIS OF THE

DISCUSSION SECTIONS IN APPLIED LINGUISTICS RESEARCH ARTICLES

WRITTEN IN ENGLISH

Evrim Eveyik-Aydın

The present study investigated the genre-specific features, namely rhetorical structure and move-bound formulaic lexico-grammatical patterns, of research article discussions written in English within the fields of Applied Linguistics and English Language Teaching (ELT). A 66,272-word specialized corpus was compiled from the discussion and conclusion sections of 36 qualitative and quantitative articles published in reputable journals. For the rhetorical investigation, Yang and Allison's structural model was used as the initial coding framework to identify communicative purposes in 2,182 sentences. The sentence-based move analysis, however, revealed that some communicative meanings are unspecified in this coding scheme, which led to the development of a more comprehensive rhetorical framework with 25 functional categories at three levels (move-step-substep). The study revealed obligatory, conventional, and optional nature of these functional categories, and the similarities and differences between the quantitative and qualitative articles both in terms of their frequency of occurrences and cyclical move-step patterns.

Once the moves were identified, the grammatical patterns were analyzed after finding key bundles in each move. Due to the small size of the data representing each move, the formulaic bundles were identified by their part-of-speech (PoS) categories. The findings provided corpus-based evidence for the existence of formulaic bundles based on four and five

PoS tag sequences in each move. Structurally, these bundles were grammatically incomplete units involving a verb phrase, prepositional phrases, and noun phrases.

The findings of the study contribute to our understanding of the writing norms of discussion sections and the development of materials that can be used in the teaching of moves, which may provide higher education students and academics with power in academic discourse and recognition by the Applied Linguistics and ELT community.

ÖZET

TARTIŞMA BÖLÜMLERİNDEKİ İŞLEVSEL GEÇİŞLER: UYGULAMALI DİLBİLİM ALANINDA İNGİLİZCE YAZILMIŞ ARAŞTIRMA MAKALELERİNİN TARTIŞMA BÖLÜMLERİNİN DERLEME DAYALI TÜR ANALİZİ

Bu çalışma, Uygulamalı Dilbilim ve İngiliz Dili Eğitimi alanlarında İngilizce yazılmış olan araştırma makalelerinin türe özgü özelliklerini, yani retorik yapısı ve işlevsel geçişlerin (moves) tekrar eden dilbilgisel-sözcüksel desenlerini incelemiştir. Öncelikle, saygın dergilerde yayımlanmış 36 nicel ve nitel araştırma makalesinin tartışma ve sonuç bölümlerinden 66,272 kelimelik özel bir derlem oluşturulmuştur. Retorik yapının incelenmesinde, Yang ve Allison'ın tartışma bölümleri için önerdiği model, çalışmanın derlemini oluşturan 2,182 cümlenin iletişimsel amacını belirlemek için ilk kodlamada kuramsal model olarak kullanılmıştır. Ancak, cümleye dayalı işlevsel geçiş analizi (move analysis), bu araçla tanımlanamayan iletişimsel anlamlar olduğunu ortaya koymuş ve bu durum üç seviyeli (ana geçişler- basamak- alt basamak) 25 farklı iletişimsel anlamı tanımlayan daha kapsamlı bir modelin geliştirilmesine sebep olmuştur. Çalışma, bu işlevsel geçişlerin, zorunlu, alışılagelmiş veya yazarın tercihine kalmış (seçmeli) olabileceğini, ve kullanım sıklığı ve oluşturdukları döngüsel desenler açısından nicel ve nitel araştırma makalelerinde benzerlikler ve farklılıklar gösterebileceğini ortaya çıkarmıştır.

Çalışma aynı zamanda, her bir işlevsel geçişe özgü tekrar eden kelime gruplarını ve dilbilgisel desenleri incelemiştir. Altı ana geçişi temsil eden datanın boyutunun sınırlı olması nedeniyle, bu anahtar kelime gruplarının bulunmasında kelimelerin ait olduğu gramer kategorileri baz alınmıştır. Çalışma her bir geçişte tekrar eden dörtlü ve beşli gramer kategorileri dizinlerinin olduğunu göstermiştir. Bu dizinlerin, dilbilgisel anlamda tamamlanmamış birimler olup eylem öbeği, ilgeç öbeği ve ad öbeği içerdiği bulunmuştur.

Çalışmanın sonuçları, tartışma bölümlerinin yazım ilkelerini anlamamıza katkıda bulunmakta ve yüksek lisans/ doktora öğrencileri ile akademisyenlerin akademik yazıda güç elde ederek alanlarında kabul görmelerini sağlamak amacıyla retorik birimlerin öğretilebilmesi için materyal üretebilmemizi mümkün kılmaktadır.

CHAPTER I

INTRODUCTION

1.1   Background to the Study

English language has spread throughout the world as 'a communicative tool of immense political, ideological, and economical power' (Kachru, 1996:910) since British colonialism in the 17th Century and the rise of United States, a former colony, as a strong military and economic power (Mauranen, 2005). Having been recognized as the language of a leading power, English has gradually become a preferred means of communication among people with different first language backgrounds within the domains of business, international relations, education, and science (Breiteneder, 2009; Seidlhofer et al, 2006). The number of its users both as a second and foreign language has increased dramatically with a 'snowball effect' as described by Myers-Scotton (2002: 34): As it is learned by more people, it has become more useful, and hence even more people have wanted to learn it. In spite of the fact that English is the fourth most commonly spoken first language of the world today, the non-native speakers of this language have outnumbered its native speakers; and its global function has been well established as the *lingua franca* of international communication (Seidlhofer, 2004).

English has gained more importance in the education system of many countries from primary to higher education since the end of the World War II (Meyer, 2004; Truchot, 2002). In some parts of Europe, for instance, learning English as a second or foreign language has been made a compulsory component of primary school curriculum replacing other most commonly taught foreign languages like German, French and Spanish. Furthermore, English has become a vehicle for the internationalization of higher education by being acknowledged

as the medium of instruction of courses and graduate programs offered in some non-English speaking countries. This position of English has been strengthened by technological advances in major English speaking countries during the post war period and most importantly by the tremendous efforts of the United States of America to support research in the newly developing areas of science. Enormous expansion in the amount of American research and publication, and the advent of computer and internet technologies that help create electronic databases to store and disseminate technical information to the world has set the new role of English as the global academic language of scientific knowledge (Meyer, 2004).

Today the dominance of English as the international language of academic research and publications is unquestionable (Flowerdew & Peacock, 2001). It is the primary language of academic events such as international conferences, panel discussions, seminars, workshops and other scholarly exchanges. Besides, the most prestigious databases, e.g. Science Citation Index (SCI) and Social Sciences Citation Index (SSCI), are of English language dominance, because a great majority of the articles in these databases come from English-language journals (Truchot, 2002). Therefore, scholars who wish to have access to such academic materials are required to learn English to follow the recent literature in their area of study, and are challenged to use standard academic English to get their work accepted to be published in these journals.

In response to the spread of English as the global language of academia, Tim Johns have coined the term English for Academic Purposes (EAP) at the University of Birmingham in 1974 (Jordan, 2002). EAP has later been established as a branch of English for Specific Purposes (ESP) along with English for Occupational Purposes (EOP). With growing number of international students who pursue tertiary studies in English and yet who need language

support to succeed, EAP has developed into a major domain in English language teaching and research to help students with their academic studies in English (Hyland, 2006).

Flowerdew and Peacock (2001) describe EAP as an international activity that is carried out in many parts of the world for a variety of purposes. It is conducted, first, in English-speaking countries where a great number of non-native speakers of English come to study in a variety of disciplines; second, in post-colonial countries where English is the medium of instruction at university level; third, in countries where English is used as a means of accessing scientific knowledge; and finally, in former Soviet-bloc countries that try to erase the influences of Russian that prevent these countries from becoming a part of global community. In these contexts EAP is concerned with a range of academic practices summarized by Hyland (2006: 1) as:

- Pre-tertiary, undergraduate and postgraduate teaching (from the design of materials to lectures and classroom tasks).

- Classroom interactions (from teacher feedback to tutorials and seminar discussions)

- Research genres (from journal articles to conference papers and grant proposals).

- Student writing (from essays to exam papers and graduate theses).

- Administrative practice (from course documents to doctoral oral defenses).

The analysis of the discourse of these academic practices and that of learners' needs are maintained as the absolute components of EAP (Flowerdew and Peacock, 2001; Johns & Dudley-Evans, 1991). In academic settings, students have been challenged to read textbooks, listen to the lectures, do presentations, get involved in group projects, and accomplish tasks with complex discourse that demands language use in different ways. Classroom interactions in which they need to get involved require them to improve their academic communication skills and abilities to read and write various texts like research articles, critical reviews,

3

technical reports, and reaction papers that are complex in terms of their content and the language used to express that content (Coxhead & Byrd, 2007; Biber, 2006). As stated by Swales (1990), each one of these is highly conventionalized genres with distinctive lexical, lexico-grammatical, and rhetorical features that show variance depending on the disciplines, tasks, and registers. As a matter of fact, it is the distinctiveness of language which makes EAP a challenging experience not only for those who use English as a second or foreign language but even for the native speakers of English with no background in academic discourse.

Thus, discourse of academic language as a critical core skill challenging all, students and scholars, regardless of their first language brings the issues of discipline-specific, and genre related academic writing needs to the attention of course providers, language teachers, and researchers. It should be noted that, the specialized knowledge of academic genres and the language required by them across the disciplines is as important for students to succeed in academic settings and for researchers to get their works published as the knowledge of the area of their study and expertise. Therefore, EAP should be relevant to discipline-and genre-specific language needs at different levels of language such as syntax, lexis, discourse, and semantics; and for that reason, the distinctive features of academic discourse should be analyzed for instructional purposes.

1.2   The Statement of the Rationale for the Study

The present study was inspired by the following issues regarding academic writing: The status of English as the language of science and technology has increased the importance of research articles (RAs) written in English as the main genre of advancing and disseminating academic knowledge among scholars all over the world. The participants of this genre, namely the researchers, are not recognized as part of discourse community unless

they share the findings of their studies with other members of the community. In order to share their findings and to contribute to the production and distribution of scientific knowledge, they need to report their studies in the form of RAs and get them published in renowned journals of the international discourse community. The knowledge of how to write a RA, then, can be acknowledged as a crucial component of research competence; because, no matter how qualified their work is, the lack of knowledge of the conventions of the specific genre of RAs leads to a lack of power in discourse, which may eventually prevent their work from getting published in journals.

Therefore, the genre analysis approach in EAP has been promoted to be able to provide researchers and students in academic settings with academic writing models that represent the conventions of RAs accepted by the discourse community. This issue has received a great deal of attention since Swales' pioneering work in which he analyzed what he called the *move-step* structure of RAs representing a variety of disciplines. Inspired by his work, many studies were conducted to analyze the organizational, also called schematic or rhetorical, structure of the conventional components (i.e. Introduction, Method, Results and Discussion) of RAs during the last two decades. Despite the abundance of such studies, the research on academic genres within the field of Applied Linguistics and English Language Teaching is still rare; and the available studies are mostly focused on the introductory sections of RAs.

Following Swales' (1990) move analysis that posited CARS model for the literature review section of RAs, many researchers have developed a particular interest in investigating the introductory sections of this genre in different disciplines to compare their findings to the moves and steps of his model. Thus, with the growing popularity of analyzing move structure of Introductions, the other parts of research articles have been left relatively less explored.

Discussion sections, however, are among the most challenging parts of research articles for many writers because it is where they interpret their findings, establish their stance by contextualizing the study in relation to the previous work in the field "reflecting a sense of membership in the larger scientific community" (Kanoksilapatham, 2005: 283).

One of the rare studies with a focus on RA discussions within the field of Applied Linguistics was conducted by Yang and Allison (2003). Their work received frequent citation due to the comprehensive seven-move model they proposed after analyzing 20 RAs that report empirical studies conducted specifically in Applied Linguistics, using the move-based approach to genre analysis. The model was commonly used as a coding framework in recent move analysis studies (e.g. Baştürkmen, 2009; Amnuai & Wannaruk, 2013); and similarly, it became a starting point for the identification of discussion moves in the present study.

However, Yang and Allison and the authors of successive studies do not mention the similarities and/or differences between the move-step structure of RAs with quantitative and qualitative approaches to research. Yang and Allison do not indicate the proportion of RAs that represent these two approaches to research in their corpus. In fact, acknowledging possible differences between the schematic structure of quantitative and qualitative RAs, either some authors limit their corpus to the quantitative studies only (e.g. Baştürkmen, 2009) or they include qualitative, quantitative and even mixed-method research reports in their corpus in quite limited numbers (e.g. Lim, 2010). Considering this as a significant gap in the literature, in order to make a more comprehensive investigation of RA discussions, the present study was inspired to include qualitative reports of empirical studies to the corpus compiled as many as the quantitative ones.

Another area that needs further exploration regarding the genre-analysis of RAs involves the investigation of the formulaic lexico-grammatical structures that are typical of

each move for pedagogic purposes. Academic writing is characterized by frequent use of formulaic language that helps authors maintain identity in a discourse community. Their absence, however, may indicate authors' inexperience in an academic context (Qin, 2014; Wray, 2002). Discipline and genre specific formulaic language can now be adequately investigated by the analysis of a representative specialized corpus using the corpus linguistic methodologies. Such corpus-based analyses of any academic genres can promote EAP instructions by providing teachers with a valuable source for developing materials to teach different aspects of academic writing, like rhetorical moves, relevant to the academic needs of scholars. The corpus-driven materials would be of great value especially for graduate students, not to mention non-natives, to cope with the demands of academic language.

In light of what is mentioned above, the present study was inspired to use corpus linguistics methodologies to investigate the organizational structure of both quantitative and qualitative RAs and their move-specific lexico-grammatical features. Thus, the study addresses a significant need for further investigation on the discussion with conclusion sections of research articles that are not treated well within this particular field.

1.3   The Purpose of the Study

The present study aims to investigate genre-specific features of the concluding sections of published RAs within the field of ELT and Applied Linguistics. The study has two purposes: a) to identify the rhetorical structure of the discussions with conclusion sections of qualitative and quantitative RAs; b) to identify the formulaic lexico-grammatical features of the rhetorical moves.

To this end, a discipline-specific corpus was compiled first in a way to include the Discussion with conclusion sections of the published quantitative and qualitative RAs selected

from the main journals in ELT and Applied Linguistics regardless of the nationality of their authors. Then, a move analysis approach was adopted for the investigation of their rhetorical structure using the model proposed by Yang and Allison (2003) as an initial coding framework. After the initial coding, the moves were redefined; and new steps and substeps were formulated following the procedure suggested by Biber, Connor and Upton (2007) for move analysis. Thus, the communicative purposes of the segments that could not be identified by Yang and Allison's model were captured in a new model proposed at the end of this part.

In the second part of the study, corpus linguistics methodologies were used to explore the lexico-grammatical features of each move identified in the first part of the study. For this purpose, the whole corpus was tagged, and analyzed by *AntConc,* concordance software developed by Laurence Anthony (2011), to see recurrent combinations in move segments. Due to the limited size of corpus data that represent each functional category, the analysis was conducted at move level to identify high-frequency formulaic bundles based on their grammatical, i.e. part-of-speech, structures.

More specifically, the present study addresses the following research questions:

1. What are the genre-specific rhetorical features of the discussion with conclusion sections of the published research articles in the field of ELT/ Applied Linguistics?

1.1. What types of moves are there in the discussion sections of the published research articles in the field of ELT/ Applied Linguistics? Which steps and substeps do the authors use to realize these moves?

1.2. What are the obligatory, conventional, and optional moves identified in the published RA discussions in the field of ELT/ Applied Linguistics?

1.3. Are there any differences between the quantitative and qualitative RA discussions in terms of move frequency? If so, what are they?

1.4. Are there any differences between the quantitative and qualitative RA discussions in terms of the move-step sequences? If so, what are they?

2. What are the lexico-grammatical features of the moves identified in the discussion with conclusion sections of the published research articles in the field of ELT/ Applied Linguistics?

2.1. Are there any strings of 5-PoS tag bundles that are typical of each discussion moves? If yes, what are these grammatical and lexical structures?

2.2. Are there any strings of 4- PoS tag bundles that are typical of each discussion moves? If yes, what are these grammatical and lexical structures?

1.4   Significance of the Study

The study explores genre-specific features of RA discussions with conclusion sections within the field of Applied Linguistics adopting Swales' move analysis approach and corpus linguistics methodologies. Taking sentence as the coding unit of move analysis, the study achieves to identify all functional units that are obligatory, conventional, and optional to use in both qualitative and quantitative discussions that comprise the corpus of the study. The study specifies some of these functional units, namely steps and substeps, as the newly emerging categories of a comprehensive rhetorical model that portrays the structural organization of RA discussions. This model that is proposed as the new coding framework for move analysis has been the main contribution of the present study.

Another contribution of the study is the identification of move-bound formulaic features for pedagogic purposes. The identification of formulaic grammatical structures and lexical expressions that are associated with each move is of great importance for the development of materials to teach rhetorical moves in academic writing courses.

1.5   Limitations of the Study

The study acknowledges that no corpus can involve all accounts of language, and that the corpus-based approaches can reveal linguistic phenomena only if the examples of those phenomena are present in that particular corpus. Therefore, although the model proposed in this study encapsulates the functions of each utterance included in the corpus of the study, more analyses should be conducted along this line with similar specialized corpora to confirm its descriptive adequacy in other RA discussions as well.

Another limitation of the study concerns the limited size of language that represents each discussion move. Although the hand-coding of each sentence may not be feasible in a larger corpus, the lexico-grammatical analyses in larger corpora may reveal more move-bound formulaic features

1.6   Definitions of Significant Terms

The operational definitions of some of the frequently used terms in the present study are as follows:

Corpus is large body of written and spoken language.

Corpus Linguistics is a sub-branch of linguistics concerned with the analysis of naturally occurring samples of written and spoken language using specially designed software and computerized techniques.

Discourse Community is a sociorhetorical network that forms in order to work towards broadly agreed sets of common goals (Swales 1990: 9).

English for Academic Purposes refers to language research and instruction that aims to understand social, cognitive, and linguistic demands of academic settings to help learners with their academic practices. In other words, it is to teach students English to facilitate their study and research in English (Flowerdew & Peacock, 2001; Hyland & Hamps-Lyons, 2002).

Genre is defined by Swales as a class of communicative events, the members of which share some set of communicative purposes that are recognized by the expert members of the discourse community (1990:58).

Genre Analysis is concerned with how language is used in a particular context for instructional purposes (Dudley-Evans, 1997). In Swales' tradition, it explores the discourse features of individual genres.

Lexical Bundles is a term coined by Biber, Johansson, Leech, Conrad and Finegan (1999: 990) to describe formulaic 'sequences of word forms that commonly go together in natural discourse' regardless of their idiomaticity. They are recurrent multi-word expressions that occur at least 10 times per million words in a genre; and are associated with certain functions in discourse.

Move Structure Analysis aims to identify the functional categories of a text in relation to typical communicative purpose of the genre (Flowerdew, 2005).

Move is a functional category that captures the purpose of a certain segment of a text. Swales (1990) defines moves as "communicative events" realized through *steps* that have specific communicative intentions. In other words, steps are rhetorical means available to the

authors to achieve certain moves. For example, *discussion* of the findings is a major move identified in the Discussion sections of RAs. The authors discuss their findings by using any one(s) of the following steps: *interpreting* the results, *explaining* the reasons for them, *comparing* them to those of other studies, and *evaluating* them.

Obligatory/Conventional/ Optional Moves & Steps: They refer to the functional categories that are obligatory, conventional, and optional to use in the representative samples of a genre. *Obligatory* moves and steps are identified in all texts within a corpus with no exception. The functions of these categories reflect the main communicative purpose of a genre or section of a genre. *Conventional* moves and steps are also functional categories with high frequency of occurrence in texts. In this study, the cut-off point was determined as occurrence in at least 60% of the corpus data based on Kanoksilapatham (2005). Those with less percentage of occurrences, however, are considered *optional* in that not every article needs to use them.

N-grams are lexical bundles that contain *n* number of words. While two-word n-grams are referred to as bi-grams and; three-word n-grams as tri-grams, and four- and five-word sequences are called four-grams and five-grams (Greaves & Warren, 2010).

PoS n-tag refers to *n* number of co-occurring tagged part-of-speech categories. 5-PoS tag bundles and 4-PoS tag bundles, which are the focus of the study, involve a combination of five and four words tagged by their PoS categories.

Schematic/Rhetorical Structure refers to the discourse/organizational structure of the texts.

1.7   Acronyms

EAP (English for Academic Purposes)

IMRD (Introduction-Method-Results-Discussion)

M (Move)

RAs (Research Articles)

S (Step)

Ss (Substeps)

CHAPTER II

REVIEW OF LITERATURE

2.1   Genre Theory and Analysis

Based on the assumption that learning occurs through discourse, genre analysis has aimed to explore the way language is used in a particular discourse for instructional purposes (Dudley-Evans, 1997). In the early 1990s, it has become the main interest of EAP research by developing in three traditions in applied linguistics identified by Hyon (1996) as English as Specific Purposes, North American New Rhetoric, and Australian systemic functional linguistics. Within these traditions, genres have been defined by their formal properties and communicative purposes with a syntactic and lexico-grammatical focus (Swales, 1990; Bhatia, 1993), by rhetorical actions they perform in recurrent situations with a pragmatic view (Miller, 1994; Berkenkotter & Huckin, 1993), and by their goal oriented social and cultural processes in certain contexts based on Halliday's view of linguistics, respectively. Each of these traditions inspired numerous genre analysis studies and had implications for language instruction with a particular focus on teaching of academic writing.

Genre theory was established within the domain of English as Specific Purposes with the pioneering works of Swales (1990) and Bhatia (1993). In his seminal book *Genre Analysis: English in Academic and Research Settings,* Swales has defined genre as

…a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rational shapes the schematic structure of the discourse and influences and

constrains choice of content and style…exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience. (1990: 58)

Genre analysis investigates the discourse features of a variety of individual genres, i.e. written and spoken text types, the forms and functions of which are defined within discourse communities that Swales has referred to as "sociorhetorical networks that form in order to work towards sets of common goals " (1990:9). According to Swales (1990: 24-27) a discourse community

- has a broadly agreed set of common public goals

- has mechanisms of intercommunication among its members

- uses its participatory mechanisms primarily to provide information and feedback

- utilizes and hence possesses one or more genres in the communicative utterance of its aims

- has acquired some specific lexis (specialized terminology, acronyms)

- has a threshold level of members with a suitable degree of relevant content and discoursal expertise.

The concepts of genre and discourse community are closely linked to each other, because it is the members of discourse community who determine the purposes of the genre and the conventions regarding its schematic/rhetorical structure.

Rhetorical structure of a particular genre includes what Swales calls *moves* and *steps* showing sequential patterning in which the text develops. A *move* refers to a segment of a text with a specific communicative function that contributes to the overall purpose of a genre by shaping its organizational pattern. Although these semantic and functional units may show variance between the different examples of a certain genre, the most common realization

pattern in a particular genre is accepted as a typical schematic or rhetorical structure by the discourse community (Swales, 1990).

The definitions of genre, discourse community and schematic/rhetorical or move structure used in the present study are based on Swales whose work has inspired many genre analysts, researchers, and EAP instructors to get focused on research article, the most frequent genre of academic writing common to all disciplines, and on its conventional components, Introduction (I), Method (M), Results (R), and Discussion (D). The rhetorical functions of these sections are commonly acknowledged in the literature: *Introduction* motivates the research; *Methods* describes the collection of data; *Results* reports the results; and *Discussion* evaluates the results (Lewin, Fine & Young, 2001:17). However, the rhetorical structure of these sections is an ongoing area of research for genre analysts.

2.1.1 Genre-Analysis Studies on Research Articles

Within the domain of academic writing, a great variety of genres ranging from PhD and MS dissertations (e.g. Bunton, 2005; Kwan, 2006; Flowerdew & Forest, 2009) to letters of application (Henry & Roseberry, 2001) has been the focus of research. However, research articles have received the most attention as the main genre of academic knowledge production. Therefore, research in genre analysis has been mostly conducted on research articles (Yang and Allison, 2003; Bunton, 2005). A review of related literature has shown that genre analysis studies on research articles have mostly investigated two aspects of this academic prose: First, the use of particular text features like hedging (e.g. Hyland, 1996), and reporting verbs (e.g. Thompson & Ye, 1991); and second, the organizational structure of RA components (e.g. Swales 1990, literature review and discussion sections; Peacock, 2002, Lim 2010 results sections; Holmes, 1997 discussion sections; Yang and Allison, 2004, results and discussion sections, Nwogu, 1997, Kanoksilapatham, 2005, all RA components) in a variety

of disciplines. These studies have significantly contributed to our understanding of the move structure of each RA section, the information presented in them and the language that is used to present this information as well as the norms recognized by discourse communities that help us produce publishable research papers.

2.1.1.1 *Move Analysis Studies on the Discussion Sections of RAs in Different Disciplines*

The move structure analysis aims to classify the segments of text in relation to the typical communicative purpose of the text (Flowerdew, 2005). This approach to genre analysis was first developed by Swales in 1981 to meet the academic needs of non-native English speakers who were engaged in research articles in English. His move structure analysis of the Introduction (I) and Discussion (D) sections of research articles in academic settings has often been cited as the most leading work on genre analysis. The move structure model that he has originally proposed for RA Introductions includes four moves: *Establishing the Field, Summarizing previous research, Preparing for Present research (often by identifying a gap in previous research),* and *Introducing Present Research.* Later, he revises this model and proposes his widely studied CARS ("Create a Research Space") Model that consists of three 'moves' accompanied by the following 'steps' based on the analysis of 48 RA introductions from a variety of disciplines including hard sciences, social sciences, life and health sciences (1990, p.141):

Move 1. Establishing a territory

Step 1. Claiming centrality   and/or

Step 2. Making the topic generalization   and/or

Step 3. Reviewing items of previous research

17

Move 2. Establishing a niche

    Step 1A. Counter-claiming   or

    Step 1B. Indicating a gap   or

    Step 1C. Question raising   or

    Step 1D. Continuing a tradition

Move 3. Occupying the niche

    Step 1A. Outlining purposes   or

    Step 1B. Announcing present research

    Step 2. Announcing principal findings

    Step 3. Indicating Research Article structure

Swales admits that Introductions are parts of research papers which many writers have difficulties with because it is the section in which the author has to decide on the amount of background knowledge to include, state the nature of the problem, establish the significance of the present research pointing to a gap in the field, and *claim centrality* for the discourse community "whereby members are asked to accept that the research about to be reported is part of a lively, significant or well established research area" (1990, p. 144). In order to establish territory, the author *makes a topic generalization* by making a general statement about *knowledge* or *practice,* or about *phenomena* (p.146), and makes a reference to the previous studies. In his analyses, Swales has found that moves and steps may be obligatory

(e.g. making reference to previous research in Move 1) or optional (e.g. announcing principle findings in Move 3); they may appear in fixed or variable order; and they may be recursive.

For the Discussion sections of RAs, Swales identified the moves of *background information, statement of results, (un)expected outcome, reference to previous research, explanation, exemplification, deduction and hypothesis* and *recommendation* (1990, pp. 170-172). Emphasizing the existence of well-established move cycles in his model, Swales (1990) holds the view that Discussion sections, in contrast to Introductions, proceed in an inside-out direction in which they move from specific information to more general interpretation of that information. In other words, Discussion sections first present the specific findings of the current study; then relate them to the existing literature, and state their general significance unlike Introductions where the works of others are primarily viewed before the present study is introduced. Therefore, Swales describes the overall shape of an empirical RA as that of an hourglass (2004:234).

One of the most comprehensive models has been proposed for RAs in natural sciences by Dubley-Evans (1994) who defines three parts in Discussion sections, namely Introduction, Evaluation and Conclusion, that involve a combination of one or two of the following nine moves: *1. information move* that provides background information about theory and research, *2. statement of result* with graphs and tables, *3. finding* without graphs and tables, *4. (un)expected outcome, 5. reference to previous research, 6. explanation, 7. claim* on the contribution of the study, *8. limitation* and *9. recommendation* for future research. In the introduction part, the writers usually follow the first three moves of the model; but in the evaluation part they use a combination of these moves depending on their purpose. For example, they follow a sequence of Moves 1, 2, 3, 4, 5, and 6 in order to *explain a result;* but when their priority is *to make a claim,* they chose move 5 or move 7 followed by move 5.

Dudley-Evans's model was later revised by Peacock (2002) who analyzed a 1.4 million word- corpus that involved a total of 252 RA discussions representing 7 disciplines - physics, biology, environmental science, business, language and linguistics, public and social administration, and law- with 36 RAs. In his corpus, Peacock has identified *claim, finding* and*. reference to previous research* as the most frequently used moves that writers in most discipline should therefore include in their discussions; and *explanation* as the least frequent move. Although Peacock has found Dudley-Evans's model quite useful for the identification of overall move structure, he has observed new move cycles in the Evaluation and Conclusion parts of discussions. For example, while two cycles (*claim* on the contribution of the study + *reference to previous work;* and *reference to previous work + claim* on the contribution of the study*)* predicted by Dudley-Evans accounted only for the 15 % of all move cycles, the unpredicted cycles of *finding + claim* on the contribution of the study*; (un)expected outcome + reference to previous research;* and *(un)expected outcome + explanation* made up 46 % of all cycles. In the evaluation part of discussions, these move cycles were observed more frequently than average in RAs from Language and Linguistics, and Law. Similarly, for the conclusion parts of discussion he has observed a frequent use of unpredicted cycles of *recommendation* for future research + *claim* on the contribution of the study; and *limitation + claim* for Conclusions.

Peacock's move analysis study has also revealed interdisciplinary and NS/NNS (Native / NonNative Speaker) variations in terms of the number of moves and move cycles used. For example, the moves of *reference to previous research*, which he observed frequently in Language and Linguistics, *limitations*, and *recommendations* were significantly less identified in Physics and Environmental Sciences. Besides, *information move* was more frequently used in Biology and Physics, but less often in Environmental Sciences and Language and Linguistics. With regard to NS/NNS variation, Peacock observed that NNS

writers used some moves less often than the NS writers (e.g. move *7. claim* in all disciplines; *8. limitation* in Physics and Biology; and move *9. recommendations* in Business, Language and Linguistics, and Public and Social Administration).

In an earlier move analysis study in Social Sciences, Holmes (1997) proposed the moves of *statement of results* and *recommendation-deduction* as the most common opening and closing moves, respectively. His analysis on the Discussion sections of 30 RAs within the fields of history, political science, and sociology revealed no completely obligatory moves, and moves that occur in cycles and in a predictable order. Holmes has also noted that history RAs differ from those of other two disciplines in terms of their overall structure with a rather long Introduction, no Method, and a brief Discussion section that do not have a cyclical structure; and therefore, History RAs, he suggests, can be considered as a subgenre of Humanities within social sciences.

Nwogu (1997) analyzed the rhetorical structure of the conventional I, M, R, D sections of 15 RAs selected from five Medical journals and identified three moves for the Discussion sections: *Highlighting overall research outcome, explaining specific research outcome,* and *stating research conclusions.* His study has shown that the authors explain research outcome by following the steps of *stating a specific outcome*, *interpreting the outcome*, *indicating significance of the outcome, contrasting present and previous outcomes* and by *indicating limitations of outcome*s. His third move of stating research conclusions, however, is realized through indicating *research implications* and *promoting further research*

In the field of Computer Science, Posteguillo (1999) used the moves identified by Swales as reference to examine RA discussions. His analyses revealed a frequent use of *statement of results* and a rare use of the move that provides *background information.* Unlike Swales' appreciation of *reference to previous research* has one of the most commonly used

moves, this move was identified only in 32% of the RAs in his corpus. Posteguillo explained this finding by the scarcity of previous work to refer to in some areas of computer science which is relatively a new area of research. In addition, regarding the move cycles, he suggested a set of cyclical patterns where *statement of results* is followed by *hypothesis,* and *recommendation.* In other words, RA writers in the field of Computer Science showed a tendency to add a comment in the form of a hypothesis or a suggestion for future research after each result.

Having observed discussion sections in 34 of the 40 RAs selected for the study, Posteguillo has also concluded that the conventional IMRD cannot be considered a typical pattern for RAs in this field.

The obligatory and optional nature of the steps in moves has also been maintained in the model proposed by Lewin, Fine and Young (2001) who observed that a move involves at least one obligatory step (head) and optional steps (pre-head and post-head). Having studied what they call the primary rhetorical functions, i.e. IMRD components, of 12 RAs in social sciences (psychology and sociology), they identified three secondary rhetorical functions (moves) for the Discussion sections: *1. report accomplishments, 2. evaluate congruence of findings to other criteria, 3. offer interpretation, 4. ward off counterclaims,* and *5. state implications.* The move analysis of their corpus has shown that Discussion sections start with reporting the accomplishments and ends with the statement of implications and requires the obligatory inclusion of either making interpretation or warding off counterclaims. The ordering of other moves (Move 2, 3, 4), however, are quite flexible. Their analysis has also shown that the Introduction and Discussion sections of the social science research articles have been characterized by different patterns of lexical choices as these sections differ in terms of the functions they have within the whole article.

The results of a move analysis conducted on all sections of 60 RAs in biochemistry by Kanoksilapatham (2005) reveals 15 Moves, four of which has been identified in the Discussion sections. In other words, in these sections, writers are observed to achieve four communicative purposes through various steps. They *contextualize the study* by *describing the established knowledge* and *presenting generalizations, claims, deductions and research gaps;* they *consolidate results* by *restating the methodology, stating selected findings, referring to previous literature, explaining differences in findings, making overt claims or generalizations* and by *exemplifying;* they *state limitations of the study* regarding *the findings, methodology, and the claims made*; and they finally *suggest further research.* Among these four moves, *contextualizing the study* and *stating the limitations of the study* have been observed in 90 % and 80% of the RAs, respectively. Therefore, these moves are considered conventional for the biochemistry RAs.

2.1.1.2   *Move Analysis Studies on the Discussion Sections of RAs in Applied Linguistics*

Within the field of Applied Linguistics, which is relatively less explored than the other disciplines, Yang and Allison (2003) analyzed 20 empirical RAs selected randomly from four established journals in Applied Linguistics (*TESOL Quarterly*, Applied Linguistics, English for Specific Purposes, and English Language Teaching Journal) with a focus on the relation between the sections of Results, Discussion, Conclusion, and Pedagogic Implications. For the Discussion sections, they identified seven moves used by the RA writers:

Move 1- Background Information

Move 2- Reporting Results

Move 3- Summarizing Results

Move 4- Commenting on Results

    Interpreting Results

    Comparing results with literature

    Accounting for results

    Evaluating results

Move 5- Summarizing the study

Move 6- Evaluating the study

    Indicating limitations

    Indicating significance/ advantage

    Evaluating methodology

Move 7- Deductions from the research

    Making suggestions

    Recommending further research

    Drawing pedagogic implication

In their paper, Yang and Allison (2003) describe the rhetorical functions of these moves as follows: *Background information move* is used to provide theoretical and methodological information regarding the study. The writers restate the aims and purposes, and the research questions or hypotheses of the study before interpreting its results. *Reporting*

results, which is the major move in the Results sections of the RAs, is used to report particular results through the presentation of statistical values, examples, graphs and tables. The summary of the integrated results, however, is presented through the move of *Summarizing Results.* The next move, *Commenting on Results,* develops in several steps in which writers *interpret* their findings and make claims based on what is suggested by them; they refer to previous research to *compare* the findings and to seek support for their explanations and claims; they give reasons to *account for (un)expected result*s; and they *evaluate results* by making comments showing personal judgment.

As for the last three moves identified by Yang and Allison, the move *summarizing the study* is used by writers to make a brief summary of the main points from the perspective of the study. Then, they *evaluate the study* by indicating its *limitations* to justify the need for further research*;* by indicating its *significance* to emphasize the contribution of the study; and by *evaluating the methodology* to justify the methods used in the study. In the final move of *deductions from the research*, writers *suggest* solutions to the problems identified by the study; *recommend* further research; and *draw pedagogic implications* by stating how language teachers can make use of the results of the study in their classes.

In this model, *Move 4-Commenting on results* was considered to be the most frequent and obligatory move, and *Move 2-Reporting results,* which had been observed in all discussion sections except for one, was recognized as a quasi-obligatory move. These two had also been observed in the Results section of RAs with significantly more frequent occurrences of the latter move. Similarly, discussion moves of *summarizing results, evaluating the study,* and *deductions from the research* were also found in the 13 of 20 RAs with a Conclusion section. Although these moves overlapped in the discussion and conclusion sections, the former gave more space to the commenting on specific results while the latter focused on the

overall results and evaluation of the study. Thus, the move analysis of the applied linguistics research articles distinguished these three sections in terms of their primary communicative purposes.

Yang and Allison's (2003) schematic structure for research article discussion moves has been the basis for some recent move-analysis studies in the field of applied linguistics. In a study that specifically aimed the investigation of the use of *Move 4-commenting on results* by expert and novice writers, Baştürkmen (2009) observed the discussion moves and steps identified by Yang and Allison to a great extent in her corpora, which included the discussion sections of 10 RAs published in Language Teacher Research Journal during 2003-2007 and the discussion sections of 10 dissertations written by MA students during the same period at a university in New Zealand. Her investigation revealed that both expert and student writers discuss their results using a repeated sequencing pattern in which *commenting on results* is preceded by *reporting of results*. Having recognized the difficulty to distinguish the move 4 steps of *'interpreting results'* and *'accounting for results'* in Yang and Allison's model, she redefined the *commenting on results* move by combining these two steps into one she preferred to name *'explaining the results'* as the term *explain* is a synonym for both *interpretation* and *accounting*.

Both groups of writers in Baştürkmen's corpora explained a result with alternative suggestions which they frequently found in literature or for which they sought support in literature. Therefore, she expanded Yang and Allison's original description of literature reference to compare the findings of current study to those of previously reported studies by showing the instances where writers referred to literature to explain results with others' views and/or to find support for their own explanations. Regarding the differences between expert and novice writers, Baştürkmen has found that the alternative explanations had been offered

significantly more often by the expert writers of RAs. This is because, she suggested, such alternative explanations were more available to the expert writers as they had better knowledge of theories and the field in general.

Yang and Allison's model was also used in a more recent study conducted by Amnuai and Wannaruk (2013) to identify the possible differences in the rhetorical move structure of discussions in research articles written in English language by Thai and international writers. Their international corpus included 30 RAs selected from 10 journals in applied linguistics and written by international writers. Thai corpus, on the other hand, included 30 RAs selected from 10 peer reviewed journals published by Thai universities and written only by Thai writers. The results of their move structure analysis in regard to move occurrence showed that *Move 4-commenting on results* was the most essential and frequent discussion move in both corpora as identified in Yang and Allison (2003) study. This was due to the fact that Move-4 encapsulates the main function of Discussion sections in research articles where the results are interpreted and related to the previous studies. The second most frequent move in both corpora has been *Move 2-reporting results*, followed by *Move 1-background information* in international corpus, but by *Move 7-deductions from research* in Thai corpus.

The international and Thai writers have shown differences in terms of their use of certain moves including the third most frequent moves in their papers. For instance, international writers have usually begun the discussion section by providing background information about the study *(Move 1)* and ended the section by commenting on results (*Move 4).* Thai writers, on the other hand, used *Move 2-reporting results* and *Move 7- deductions* as the most frequent opening and closing moves, respectively. In general Thai writers seemed to be more concerned with drawing pedagogic inferences and making suggestions for future studies based on the findings of their study. Their tendency to use more deduction steps is

explained by Thai researchers' particular interest in reading this section of research articles to get practical suggestions on possible future research topics in this promising field of English language learning and teaching. International writers, on the other hand, used Move 6 more often than Thai writers to evaluate their study by indicating its limitations and significance.

Identification of the rhetorical structure allowed in a particular genre with the obligatory and optional moves, and of their key linguistic features has resulted in the development of genre-based pedagogies which provide language learners with an explicit teaching of the way texts are structured within a target genre. These pedagogies have been inspired by the assumption that in order to teach a genre effectively, language teachers themselves need to be aware of the communicative purpose of each move and the linguistic features associated with them. Despite the abundance of early genre-based studies analyzing the schematic structures of research articles in a variety of disciplines, the linguistic realizations of each move is reported to be relatively less investigated mainly due to the methodological difficulty of determining the relation between rhetorical function and its linguistic form. As discussed by Flowerdew and Forest (2009), a thorough investigation of such relation is a time-consuming work that requires a systematic examination of large amounts of data, which has only become possible with the advent of *corpus linguistics* and the integration of its methods into genre analysis studies.

2.2 Corpus Linguistics in Genre Analysis

'Corpus' is a Latin word for 'body'. In linguistic terms, it refers to a large body of authentic written and spoken language; and hence, corpus linguistics is simply concerned with the analysis of these naturally occurring samples of language. Despite the tendency to consider it a new branch of linguistics or a theory of language, some researchers like Granger

(2002) and McEnery & Wilson (1996) defined corpus linguistics as a methodology that could be used in many areas of linguistic enquiry like lexis, syntax, or semantics.

Indeed, studies in corpus linguistics have been conducted ever since *Brown Corpus of Written American English* and *Lancaster-Oslo-Bergen (LOB) Corpus,* the first language corpora with one million-word each representing American and British English, were compiled in the early 1960s. However, it began to flourish as a methodology for language analysis in the 1980s. During this period, computers were introduced to the field of linguistics as tools of storing, organizing, annotating and searching vast amount of language corpora. Thanks to fast technological advances, enormous collections of language texts became electronically available and easy to analyze without requiring computational expertise. As a result, the studies which previously took considerable amount of time and effort were accomplished easily, rapidly and with higher reliability through the corpus linguistics software specifically designed for the analysis of large amounts of electronic corpora.

The software packages used by corpus linguistics (e.g. *WordSmith Tools* developed by Mike Scott, and *AntConc* by Laurence Anthony) offer tools allowing both quantitative and qualitative analyses. Such computerized text analysis programs, also known as concordancing programs, are used to access texts of a certain genre for a variety of purposes: Word lists can be generated based on frequency count calculations; *key words* that are unusually frequent in certain texts or part of a text can be determined; every occurrence of such words and phrases in corpora can be sorted and displayed within their immediate linguistic contexts; *cluster* analysis can be conducted to show the recurring sequences of word combinations, and their structural forms can be identified and thus, lexico-grammatical profile of the language used in a certain genre can be obtained (Baker, 2006; Scott & Tribble, 2006).

Today, these innovative techniques brought by computerized text analysis programs have given a new direction to language studies. The role of corpora was established as sources for compiling dictionaries and grammar reference books, e.g. Sinclair's pioneering COBUILD project (Sinclair 2004), and *Longman Grammar of Spoken and Written English* (Biber 2006); for creating frequency-based word lists in academic discourse, e.g. Coxheads' (2000) Academic Vocabulary List; and for analyzing the distinctive linguistic features of a variety of genres at different levels, e.g. the use of linking adverbials (Conrad, 2004). The importance of corpora as valuable sources for language teachers as well has been acknowledged by many authors including Kennedy (1998), Conrad (1999), and Tribble (2001) who elaborated on its role in the selection of items to be included in course syllabus.

Even more, the relevance of the use of corpus techniques in the analysis of genres was acknowledged by the pioneer of genre studies, Swales himself, who described 1990s as the *infancy* period of corpus linguistics (2004: 8). During those years when he was composing his *Genre Analysis*, the problem, he complained, was the lack of corpus materials compiled for those working in EAP and ERP (English for Research Purposes). Today, on the other hand, despite having all these corpus techniques within our reach, the main problem has to do with finding the best ways to incorporate them into our research. At this point, finding or compiling the most appropriate corpora in terms of type and size depending on the research purpose poses a serious challenge for the researchers.

2.2.1 Specialized Corpora

For the last decade, the size and variety of corpora have increased dramatically gaining worldwide popularity (Johansson, 2008; McEnery, Xiao and Tono, 2006). Among many corpora compiled over the years for various purposes were corpora of historical texts, of different language varieties and disciplines, of special text types, and of English for Specific

Purposes. These corpora came in a variety of sizes from "large" (e.g. *British National Corpus* representing British English) to "small" and to "specialized" (e.g. corpus of medical research articles) depending on their contents and research purposes for which they were created (Gavioli, 2005). Many linguists have referred to them either to find empirical evidence for their hypotheses on each linguistic level, or to employ corpus tools for the quantitative language analysis in many areas of linguistics (Lüdeling and Kytö, 2008).

General corpora help us make generalizations about the way language operates as a whole (Flowerdew, 2004), therefore it is used for compiling dictionaries, grammar books and textbooks. Specialized corpora, on the other hand, provide insights into the use of language in a particular text type such as newspaper editorials, academic research papers, casual conversations and so on. In other words, specialized corpora represent the language of a particular domain selected for a specific purpose. As stated by Hunston (2002), although the degree of specialization is not restricted, there are some parameters used to define specialized corpora. Flowerdew (2004: 21) summarizes these parameters as follows with the illustrative examples:

Table 1. Parameters for Defining Corpora as Specialized

Parameters	Details/Examples
Specific purpose for compilation:	To investigate particular grammatical, lexical, lexico-grammatical, discoursal or rhetorical features
Contextualization:	Setting (e.g. lecture hall)
	Participants (role of speaker/listener; writer/reader
	Communication purpose (e.g. promote, instruct)
Size:	
Whole corpus	1.5 million words
Sub-corpus or small scale corpus	20,000-250,000 words
Genre:	Promotional (grant proposals, sales letters)
Type of text/ discourse:	Biology textbooks, casual conversations
Subject matter/topic:	Economics, the weather
Variety of English:	Learner, non-standard (e.g. Indian, Singaporean)

Source: Flowerdew (2004:21)

2.3   The Use of Corpus Linguistics in the Investigation of the Lexico-Grammatical Features of Academic Discourse: In Search of Formulaicity

Corpus linguistics methodologies allowed researchers to make comprehensive studies that aimed to find defining linguistic features of written and spoken genres (Biber, 2006). The most outstanding and inclusive example of such studies has been the *Longman Grammar of Spoken and Written English* (LGSWE), a corpus-driven reference grammar of English, that provides distinctive structural characteristics of academic discourse, conversation, fiction and newspapers (Biber, Johansson, Leech, Conrad, & Finegan, 1999). The academic discourse

subcorpus of the *Longman Spoken and Written English* (LSWE) Corpus included more than 5-million words representing both research articles (2.68 million words) and academic books from a variety of disciplines. The results of the comparative analyses revealed three grammatical classes that are more common in academic language than in conversation, fiction and newspapers (Biber, 2006: 14-17):

- Nouns (300,000 nouns per million words in academic writing vs. 150,000 per million words in conversation): especially those formed with *–tion* and *–ity*, noun phrases with multiple post-modifiers (e.g. *the utilization of such devices for social purposes*), and stance noun+ of-phrase (e.g. *possibility of*, *value of*, *importance of* etc.)

- Adjectives (80,000 adjectives per million words in academic writing vs. 20,000 per million words in conversation): derived adjectives, especially those derived with *–al* like *functional*; predicative adjectives (e.g. *different*, *important*, *difficult*, *necessary*, *possible, necessary, useful* etc.); attributive adjectives (e.g. *the basic logical content*)

- Prepositions and prepositional phrases as post-modifiers in noun phrases (e.g. the effect on the final state)

Abundant use of nouns and complex noun phrases followed by prepositional phrases as the most outstanding feature of academic writing was also emphasized by Coxhead and Byrd (2007) who defined academic language as noun-centric. However, the LGSWE identified the characteristic uses of other grammatical classes in academic language. Although verbs, adverbs and adverbials, for instance, were observed less often in academic discourse in comparison to the other three registers, the following categories emerged as typical of this type of language (Biber, 2006: 16-17):

- copula be, copular verb become, derived verbs (especially those derived with *re-* and *–ize*); existence verbs (*include, indicate, involve* etc.), activity verbs (*use, produce,*

provide, apply, form, obtain, reduce); communication verbs (*describe, suggest*); mental verbs (*consider, assume, determine*)

- some prepositional verbs: activity verbs (*deal with, be used in, be applied to, be derived from*), communication verbs (*refer to*); mental verbs (*be known as*); causative/existence/occurrence verbs (*lead to, result in, occur in, depend on, consist of, be based on, be associated with, be related to*)

- specific categories of adverbs and adverbials like linking adverbials (*thus, however, therefore, for example*), purpose and concessive adverbials (*in order to, although*)

- some specific verbs in passive voice (*be+ made, given, taken, used, found, seen, shown, considered*) with no *-by phrase*

In addition to all these, the use of – *that* and *–to clauses* controlled by stance adjectives (e.g. *It's important/ (im)possible/difficult, hard, necessary/(un)likely that/to …)* also emerged as the most distinctive features of academic language.

The LSWE Corpus provided evidence for the formulaic nature of academic language as well. Corpus linguistic methodologies used in this project played a significant role in the identification of language formulaicity as acknowledged by authors like Simpson-Vlach & Ellis (2010) and McEnery & Wilson (1996). The computational analyses conducted through corpus tools enabled Biber et al. (1999) to document the frequently used patterns of multi-word combinations that they called *lexical bundles*. This term was later adopted by numerous researchers in the succeeding studies on formulaic language.

2.3.1 Identification of Formulaic Language and its Functions

Language is *formulaic* on the grounds that it is composed of recurrent word combinations that are available to the speakers and writers of a language within the format of

"semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments"(Sinclair, 1991: 110). For the last two decades, what Sinclair called 'semi-preconstructed phrases' in his 'idiom principle' has been given different names like *lexical phrases* or *prefabricated language chunks* (Nattinger & DeCarrico, 1992), *formulaic sequences* (Wray & Perkins, 2000), *morpheme equivalent units* (Wray, 2012), and *lexical bundles* (Biber et al. 1999); and defined to encompass a wide range of expressions including *idioms* with fixed lexical items (e.g. *spill the beans), lexical collocations* (e.g. *curry flavor*), and *grammatical collocations* (e.g. *the ADJ-er the ADJ-er*).

Definitions of multi-word combinations reflected different approaches to identify the formulaic expressions and the functions they accomplish in written and spoken discourse. Having considered the mental processing involved in formulaicity, Wray and Perkins (2000), defined formulas as prefabricated phrases that are "stored and retrieved whole from memory at the time of use" without being analyzed structurally by the language grammar (p.1). Analytical processing, however, is called for only when novel structures are encountered as Wray puts in her "needs-only-analysis" principle (2008). Although the holistic processing of language seems to be preventing novelty and creativity in language production, it offers significant advantages to language users. First, storing and processing formulaic phrases as single units increases fluency by reducing the mental stress and the time spent on demanding tasks. Second, they compensate memory limitations by functioning as time-buyers while speakers plan their responses in conversations (Wray & Perkins, 2000; Wray, 2008). For a successful communicative language processing, however, Wray and Perkins (2000) emphasize the necessity of establishing a balance between holistic and analytical processing through which language users retrieve the formulaic structures effortlessly and analyze what is novel from scratch.

Wray and Perkins (2000) and Wray (2012) also elaborated on the role of formulaic language in our social interactions. They identified three communicative functions achieved by formulaic sequences: First, language users *manipulate the world* around them using the formulas of commands (e.g. *hand it over*), requests (e.g. *could you repeat that please?*), politeness (e.g. *I wonder if you'd mind…*) and so on. Second, they *assert their separate identity* using phrases of personal turns (e.g. *I wanna tell you a story*), story-telling (e.g. *You're never going to believe this, but ….*) and the like. Finally, they *assert group identity* using forms of address (e.g. *your highness*), group chants (e.g. *we are the champions*) institutionalized words (e.g. *happy birthday*) and so on. In a way, language users rely on these formulaic sequences in their interactions in order to meet their physical and social survival needs.

Another approach used in the identification of formulaic language is frequency-driven. Biber and his colleagues have often adopted this approach in their studies using corpus linguistic tools (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Biber, Conrad & Cortes, 2004; Biber & Barbieri, 2007; Biber, 2009). In the1999 project of LGSWE mentioned earlier, they coined the term *lexical bundles* for the 3-word, 4-word- and 5-word sequences, i.e. *n-grams*, that occur more often than expected by chance. Their cut-off frequency criterion to qualify a lexical bundle in this study was at least 10 times of occurrence per million words, in at least 5 different texts of a certain genre to avoid idiosyncratic uses of individual language users. In the 2004 study, however, Biber, Conrad & Cortes set a more conservative frequency cut-off of 40 times per million words.

In support of this approach, they argue that high frequency is a reflection of formulaicity. It reveals the formulaic patterns that otherwise would not be noticed by the researchers, because most of them are not idiomatic in meaning, and therefore, not as salient

as fixed idiomatic expressions. Neither are they complete structural units (e.g. *I don't know what*). In fact, Biber et al. (1999: 995) found that only 15% of lexical bundles identified in conversation and 5% of them in academic prose are complete structural units. In their corpus data, the bundles with incomplete structures usually connected two *clauses* (e.g. *I want to know*) in conversation, and two *phrases* like a noun phrase + beginning of a prepositional phrase (e.g. *the end of the, as a result of)* in academic prose. Such bundles with a noun phrase or a prepositional phrase comprised 60% of all bundles in this register.

Biber et al. (1999) accept that the patterns identified this way are not explanatory. On the contrary, they believe frequency-based methodologies reveal patterns that must be explained, especially in relation to their discourse functions. Biber, Conrad & Cortes (2004) grouped the lexical bundles into taxonomy with three main categories according to their functions: *Stance bundles, discourse organizers,* and *referential expressions.* S*tance bundles* are expressions of certainty or uncertainty (*epistemic stance bundles*), and of attitudes, obligation/directive, desire, and intentions/predictions (*attitudinal/ modality stance bundles*). The examples include bundles like *I don't know if, are more likely to, might want to,* and *you need to know. Discourse Organizers* are used to introduce a new topic (*topic/introduction focus bundles*), and to provide explanations about a topic (*topic elaboration/ clarification bundles*). For example, the bundles like *want to talk about,* and *what I want to do* were mainly observed in classroom teaching to introduce a new topic, while *you know I mean, on the other hand, as well as the* were used to provide additional explanations, and to make explicit comparison and contrast. *Referential bundles*, on the other hand, serve to identify an entity (e.g. *those of you who*), summarize the main points after a long explanation (e.g. and *that's one of the*), and initiate a discussion by stating the focus point (e.g. *one of the things*). They also indicate imprecise reference (e.g. *something like that);* specify the attributes of the following head noun (e.g. *a little bit of*); and specify times, places, or locations in the text (e.g.

As shown in Figure 5). Biber, Conrad and Cortes (2004) found that most of the bundles identified in academic writing had referential functions that were primarily realized by noun phrase and prepositional phrase fragments. Most stance bundles had dependent clause fragments, while all of the stance prediction bundles were composed of verb phrase fragments. Discourse organizers, however, included both verb-based bundles, dependent clause bundles, and noun/prepositional phrase bundles. These findings revealed the structural associations of lexical bundles within particular registers.

Identifying formulas by only extracting word sequences with high-frequencies of occurrence, and then assigning them to certain functions is not without limitations, though. Wray (2002) points to the fact that this approach may fail to reveal many formulaic phrases or idioms like *long live the king* and *kick the bucket* that are most likely to occur below the cut-off frequency threshold even in a large corpora, because the messages they convey are quite rare. Therefore, she suggests that rather than relying solely on the frequencies of occurrence, the ratio of message-expression, that is the extent to which certain forms are preferred over the others to express a particular message, should be considered to define formulas. In other words, the context in which these linguistic forms are uttered, and their functional meanings must be part of the identification from the very beginning of this process.

2.3.1.1 *Corpus-based Bundle Studies in Written Academic Discourse*

Despite its limitations, numerous studies have been conducted to empirically investigate the structural features and functional uses of 3- word, 4-word, and 5-word bundles in academic writing drawing on Biber and his colleagues' frequency-based approach. In these studies, the researchers were primarily concerned with the role of disciplinary variations, language- and genre-specific features, nativeness (native versus non-native writers), and of experience (expert versus novice writers) in the formulaic use of academic language. Some of

these studies will be reviewed below to show the range of research interests regarding the lexical bundles.

One of the studies with a focus on language-specific variations was conducted by Cortes (2008). She analyzed the corpora of History articles published in American journals in English and in Argentinian journals in Spanish in order to identify the most frequently used four-word lexical bundles. Her aim was to investigate structural and functional similarities and differences in their use across languages. For this purpose, she first identified them by means of a specially designed software called Lexical Bundles Program (LBP); and then analyzed their discourse functions based on the taxonomy offered by Biber, Conrad and Cortes (2004). Her findings confirmed the previous findings that most of the lexical bundles in academic writing were phrasal; mainly involving prepositional and noun phrases. The analyses of the two corpora revealed that Spanish History writing had twice as many bundles as English History writing owing to the various linguistics features of Spanish in which some grammatical classes like nouns, pronouns, and demonstratives are marked for number and gender. Spanish nouns are rarely pre-modified; instead, they are post-modified by using adjectives and prepositional phrases. Noun-noun pre-modification, for instance, which is quite common in English (e.g. *immigration history*) is not allowed in Spanish but realized through post-modifying prepositional phrases (e.g. *la historia de la immigración*). Therefore, she explained, Spanish corpora of History writing included many *noun phrase + prepositional phrase fragments*, or *prepositional phrase + prepositional phrase* fragments. Besides, she also explained that some four-word lexical bundles in Spanish may be expressed by shorter bundles in English as in the example of 3-word bundle *of human rights* versus 4-word bundle *de los derechos humanos* where Spanish plural forms carry a determiner unlike English. Bundles with equivalent expressions, which are the direct translation of each other, made up of 21 % of all bundles identified in the study.

However, English corpus was observed to include a wider variety of prepositions than those in Spanish corpus. English prepositional phrases were introduced with the prepositions of *after as, at, by, for, from, in, of, on, over, to* and *with*, while fewer number of prepositions were found in this position in Spanish. This little one-to-one correspondence between the prepositions in both languages, Cortes argued, may cause difficulties for English and Spanish speakers learning these languages.

Cortes also found similarities in the functions of the identified bundles in both corpora. Most of them were referential expressing time-event (e.g. *in the nineteenth century, after World War II*), place-event (e.g. *in the United States, the civil right movements*), multifunction (e.g. *at the turn of, the end of the)*, quantifiers (e.g. *for the first time)*, identification (e.g. *one of the most*), and framing attributes (e.g. *in the context of, from the perspective of, in the hands of).* She found only a few stance epistemic personal bundles (e.g. *the fact that the)* and discourse organizers (e.g. *on the other hand, as well as the).* She observed that the discourse organizers in English were used for the purposes of topic elaboration and clarification while those in Spanish served to introduce topic.

In an earlier study, Cortes (2004) investigated the structural and functional uses of four-word bundles by the authors of published history and biology RAs and by the student writers at different academic levels within these disciplines. Her corpus of published RAs included 92 texts in history and 199 texts in biology written by different authors. The corpus of student writing, however, included a total of 277 (233 undergraduate level and 44 graduate level) research papers and assignments written by students majoring history; and 213 texts (162 undergraduate level and 51 graduate level) written by those majoring biology. In the corpus of published history writing, the study identified 54 target bundles with referential functions indicating time, place and quantities, or used as text-organizers. Cortes who used Biber's taxonomy for functional analyses, identified two new groups of referential bundles

that she defined as quantifying (e.g. *a member of the, as part of the, the majority of the)* and subject-bound (e.g. *in the history of, the history of the*). Text organizers, on the other hand, indicated comparison /contrast (e.g. *on the other hand, in the first place*), inference (e.g. *as a result of, on the basis of*), and framing conditions used to interpret discourse (e.g. *from the perspective of, the extent to which, the fact that the*).

In the corpus of published biology writing, on the other hand, the study identified 109 target bundles, which showed structural and functional differences. While history bundles only included *noun* and *prepositional phrases*, biology bundles belonged to a wide variety of structural groups like *noun phrases*, *prepositional phrases*, *it + Verb be+ adjective clauses, Verb be + complement clauses, noun phrase + Verb + complement clauses*, and *passive* constructions followed by *that-clauses* (e.g. *has been shown that*). Besides, stance bundles showing probability in biology (e.g. *is likely to be, it is possible that, the probability that the*) were not frequently used by the authors of the history writing. To be able to classify bundles functionally, Cortes needed to create new categories for biology as well. She named these categories referential quantifying bundles (e.g. *a large number of, a measure of*), referential quantifying statistical bundles (e.g. *was positively correlated with, not significantly different from*), referential descriptive bundles (e.g. *the depth of the, the length of the*), other stance bundles (e.g. *not appear to be*) and other bundles in biology (e.g. *in the evolution of the, the genetic basis of*).

As for the differences between expert (writers of published RAs) and student writers, the study revealed bundles that were never used (e.g. *from the perspective of, on the eve of,* and *the eve of the*) or rarely used (*in the course of, in the context of, the ways in which, the extent to which, to the extent that, to mention just a few, at the turn of,* and *in the wake of*) by the student writers of history. Similarly, the students of biology never or rarely used some

bundles (e.g. referential bundles like *the shape of the, with the number of;* text organizers like *the extent to which, with respect to)* identified as frequent in academic prose in this discipline. The students at higher academic levels, however, used more of target bundles than lower level students in biology. Whereas, the history students at graduate level did not differ from the lower level students with respect to the frequency in the use of bundles despite their more frequent exposure to published articles during their graduate studies. Besides, some bundles used by lower level students did not always serve the same functions as those conveyed in published articles and in upper level students' papers in both disciplines. Considering these findings Cortes concluded that exposure to published academic materials alone may not be sufficient for students to use the bundles of professional prose with functions intended by the expert writers. Therefore, students' awareness of the structural and functional uses of lexical bundles should be raised by explicit teaching of them.

In a recent study conducted by Qin (2014) with a different perspective, five-unit bundles in 128 RAs published in Applied Linguistics were explored first in order to compile a reference corpus for the investigation of the use of these constructions by non-native graduate students and by native-English speaking academics. In her study, she was inspired to find whether "L2 graduate student writers' use of lexical bundles increases in line with their levels of study to approximate to expert writers' usage" (p.221). With this purpose, her analysis corpus included 136 academic papers written by 20 non-native MA and PhD students with different nationalities in an American University, and 15 published RAs written by 11 native English speaking professors of Applied Linguistics. The analysis of her reference corpus revealed that 54% of the bundles identified in published RAs was made up of noun phrases and prepositional phrases while 46 % of them were verb-based phrases or clauses. Noun phrases generally included *of*-phrase fragments (e.g. *the results of this study*) or other post-modifier fragments (e.g. *the fact that the, English as a second language);* and prepositional

phrases included embedded *of*-phrases *(e.g.in the case of)* and other prepositional phrase fragments (e.g. *at the same time*). Verb–based phrases, on the other hand, mostly included structures like *anticipatory it + verb/adjective phrase* (e.g. *it is important to, it has been argued that*), *verb/adjective + to-clause fragment* (e.g. *does not appear to, is interesting to note that*), *copular be + noun/adjective phrase* (e.g. *is one of the most, this is consistent with*) among others. As for the functional meanings revealed by these structures, she proposed 4 new sub-categories to classify referential bundles: *brevity bundles* to omit some information to prevent redundancy (e.g. *and so on), explanation bundles* to provide explanations (e.g. *that is to say*), *exemplification bundles* used for the purpose of illustration (e.g. *for example, the*), and *locator bundles* to build link between the present and previous discourse (e.g. *as noted earlier*).

Another finding was that as students' academic study level increased, the number of bundles identified in their papers also increased steadily. PhD students who spent more than 2 years in the program used more bundles than the 1st and 2nd year MA and PhD students past the second year. The bundles they used even outnumbered those used by the experts, but these bundles represented fewer types of bundles, which showed their tendency to use certain constructions in their writings repeatedly. In the meantime, academic structures *like noun phrases with post-modifier fragments* including *past participles and prepositional phrases* were used more often by graduate writers at the higher levels of study than lower levels. This finding suggests that the lower level graduate writers need to be taught how to express information more economically by using more complex and target-like structures.

PhD students of the study were also found to use more *text organizer bundles* with the purposes of explaining, exemplifying and providing focus, and *stance bundles* than MA level students. Qin explains this finding by the higher requirements regarding the discourse quality

of the written assignments at PhD level, and by PhD students' increasing ability to view their writing from the perspective of their readers.

The use of lexical bundles in native speaker writing versus non-native speaker writing has also been the interest of research. To test the hypothesis non-native users of language have a more limited repertoire of lexical bundles overall, Adel and Erman (2012) investigated the corpus of Stockholm University Student English Corpus (SUSEC) that included 325 essays written by native speakers of Swedish and of British English majoring linguistics at undergraduate level. Their findings confirmed the hypothesis in that the native students of English produced a considerably wider range of lexical bundles of different types in their writings than the non-native students of English. The study identified 130 bundles that were unique to the former group and 60 bundles unique to the latter group. Both groups of writers shared a total of 55 bundles (22 %). Similarly, Chen and Baker (2010) who compared Chinese EFL university students' writing to the writings of native English-speaking university students and native experts found that Chinese students used the fewest number of lexical bundles with a tendency to overuse certain ones.

The significance of the studies on lexical bundles, a couple of which were mentioned above, is that they established a strong structure-function association; identified the recurrent word patterns overused and underused by the native versus non-native, and student versus expert writers' of the academic prose; and revealed the variations specific to the disciplines, languages and genres. Thus, they shed light on our understanding of how certain word combinations are associated with certain communicative functions; and to what extent they are used by those involved in academic writing. What is more significant is that they provided evidence for the formulaicity of academic language and drew attention to their pedagogic teaching to maintain an identity in a discourse community.

2.3.1.2 Lexical Bundles and Rhetorical Moves

Despite the considerable number of studies that identified lexical bundles, very few attempts were made to search formulaicity in rhetorical moves. A few of the studies that primarily aimed to identify rhetorical move-step structure of RAs looked for the signaling expressions and their linguistic features that seemed typical of certain moves (e.g. Kanoksilapatham, 2003; Lim 2010). Yet, their findings were limited on the grounds that they could not empirically relate these expressions and their linguistic features to the moves.

Inspired by the fact that "once identified, these linguistic features could help better describe and illustrate the communicative functions of each move" (p.34) to teach a genre more effectively, Cortes (2013) analyzed a corpus of RA introductions to find four-word and longer bundles. The corpus of the study included 1,372 RA introductions from 13 different disciplines including applied linguistics represented with 87 texts. Her analysis revealed 135 lexical bundle types (3,849 tokens) that included combinations of four-to-nine words (*the rest of the paper is organized as follows, the reminder of the paper is organized as follows).* She commented that nine-word lexical bundles had never been reported before in previous studies. As confirmed many times by the findings of previous studies, however, the identified bundles had noun phrases or prepositional phrase fragments (*in the present study, the objective of this paper),* verb phrase fragments (*is related to the, it has been shown that),* and dependent clause fragments (*that there is a, we show that the*).

Once the bundles were identified, she conducted a move analysis based on Swales' CARS model for RA introductions to find the moves/steps in which they occurred. In her paper, she provided the list of bundles frequently used in segments functionally defined by CARS moves and steps.   To illustrate a few, bundles like *in the field of, in the absence/presence of,* and *to the use of* were linked to Move 1, Step 3 (Establishing a territory, reviewing items of previous literature); while bundles like *an analysis of the, in the context of*

the, and *the ways in which* were linked to Move 3, Step 2 (Presenting the present work with its research questions and hypothesis). She observed that bundles, especially longer ones, triggered the communicative function of the move/step in which they occurred. Shorter ones (4-word and 5- word bundles), however, complemented the language that signified the beginning of a move.

Another study that is worth mentioning due to its methodology was conducted by Durrant & Mathews- Aydınlı (2011) to identify recurrent bundles in a comparative analysis of the introduction sections of the student essays and published RAs. The essay corpus of the study included 94 essays written by MA students in social sciences at British Universities. These texts comprised a subset of the British American Written English Corpus. The article corpus, on the other hand, comprised 94 journal articles. Drawing on function-first approach suggested by Wray (2002), they first tagged the text by their communicative functions considering the move types defined in literature, and then identified the recurrent formulas associated with these functions. Their focus was on the specific communicative function of Move 3 (*Essay focus*), Step 3 (*indicating structure*) through which the structure of essay was described by indicating *what will happen in the text and where in the text it will happen* (e.g. *In the final section we discuss implications of this research for…*). They found that *what* function of Move 3 Step 3 is conveyed by structures with *text(subject)+ verb, passive voice,* and *pronoun + verb*, while *where* function is realized by *text + verb, adverbials*, and *verb markers* . Some parts of the forms identified were highly formulaic. Their comparative analysis revealed that this function is more commonly used in student essays than in RAs in which its use shows disciplinary variations.

In light of all these, in order to provide academics and higher education students with power in academic writing, the present study was inspired to investigate genre-specific features of RA discussions within the field of ELT/ Applied Linguistics with a focus on their

rhetorical structures and the neglected move-bound lexico-grammatical features. The following section describes the methodology of the study in detail.

CHAPTER III

METHODOLOGY

3.1   Introduction

This chapter includes the methodology of the study. It describes in detail the procedure followed to construct a specialized corpus to address the research questions of the study and to do move structure analysis that led to the formulation of a coding scheme with new categories at step level.

The study aims to investigate the rhetorical structure and the formulaic lexico-grammatical features of the Discussion with conclusion sections of the RAs published in the field of ELT/ Applied Linguistics. More specifically, the following research questions are addressed:

1. What are the genre-specific rhetorical features of the discussion with conclusion sections of the published research articles in the field of ELT/ Applied Linguistics?

1.1. What types of moves are there in the discussion sections of the published research articles in the field of ELT/ Applied Linguistics? Which steps and substeps do the authors use to realize these moves?

1.2. What are the obligatory, conventional, and optional moves identified in the published RA discussions in the field of ELT/ Applied Linguistics?

1.3. Are there any differences between the quantitative and qualitative RA discussions in terms of move frequency? If so, what are they?

1.4. Are there any differences between the quantitative and qualitative RA discussions in terms of the move-step sequences? If so, what are they?

2. What are the lexico-grammatical features of the moves identified in the discussion with conclusion sections of the published research articles in the field of ELT/ Applied Linguistics?

2.1. Are there any strings of 5-PoS tag bundles that are typical of each discussion moves? If yes, what are these grammatical and lexical structures?

2.2. Are there any strings of 4-PoS tag bundles that are typical of each discussion moves? If yes, what are these grammatical and lexical structures?

3.2 Data

Data is a 66,272- word genre specific corpus that was compiled electronically by considering the theoretical issues discussed by the authors who did considerable work in corpus linguistics (Swales, 1990; Flowerdew, 2004; Biber, Connor, and Upton, 2007). The corpus data included the discussion sections of 36 published RAs that report empirical studies in the field of English Language Teaching and Applied Linguistics.

The corpus of the study was small, and specialized. Justification for the use of small size specialized corpus was found in the writings of several authors like Flowerdew and Forest (2009), Fuertes-Olivera (2008) and Ghadessy et. al. (2001). They suggested that the corpus that includes the texts of the same genre and discipline may produce sufficient data for the purpose of the analysis regardless of their size. Limiting corpus to a specific genre within a particular discipline also controls possible disciplinary variations (Kanoksilapatham, 2005).

Besides, a small corpus enables some analyses that require the hand-coding of moves which otherwise cannot be handled manually within a large corpus.

"Representativeness" is an issue that caused corpus linguistics to be criticized highly on the grounds that a corpus includes only a small sample of language that cannot be the representative of a large population (McEnery and Wilson, 1996). In order to achieve representativeness in the present study, the corpus was built in a way to include the samples of the same genre belonging to the same field, with a specific purpose and with sections that are specifically introduced with the heading 'Discussion' and 'Conclusion' as suggested by Biber et. al (2007).

The corpus of the study was constructed and analyzed as explained in the following sections.

3.2.1   Compiling the Corpus Data

The research articles selected to compile the corpus were published in the field of Applied Linguistics during the period 2007-2012. Since the corpus was meant to represent the recent academic language used in publications and the recent discussion moves accepted by the discourse community, those published in earlier years were not considered for the study (See Appendix A for the list of research articles included in this corpus).

All of the RAs were written in English by the authors of different nationalities. Research has shown the underuse or overuse of some moves and move cycles in the non-native corpus (e.g. the underuse of *making claims* in Peacock's 2002 study; the overuse of *deduction move* by Thai writers in Amnuai & Wannaruk's 2013 study). However, the articles were not selected on the basis of authors' first language background, because the purpose of the study was to explore how authors discuss their results and conclude their papers within the

framework of the established norms, and to reveal possible rhetorical differences between the reports of quantitative and qualitative empirical studies, not between the language used by native and non-native writers of English. Besides, since these articles had already been published by the respected journals, their authors were assumed to be proficient users of academic English confirming to the norms regardless of their first language, and therefore, "nativeness" was ignored as a selection criterion.

The selection of the research articles, however, was based on purposive sampling to include the qualitative and quantitative reports of empirical studies with separate discussion and conclusion sections. As research articles may show variance in terms of the wording of their conventional section headings, electronically accessed journal issues were screened to find those with separate discussion and conclusion sections. This is an important criterion because each section (I, M, R and D) of RAs has a different function in relation to the overall purpose of this genre, and is characterized with certain moves as shown by move analysis studies since Swales' work in 1990. Thus, articles with combined results and discussion sections, for instance, were excluded to be able to identify the moves that were particularly specific to the concluding sections of an empirical study.

It should be noted that of the selected quantitative RAs, seven were written by one author, seven by two authors, and four by three authors; and of qualitative RAs, 13 were written by one author, four by two authors, and one by one author. Although articles written by the same authors were not included in the corpus in order to avoid idiosyncratic uses of language, those with co-authors were not left out of consideration, because the study aimed to identify possible moves and steps used in the discussion of research findings. In other words, the study did not aim to compare the rhetorical structure of RAs with single author to that of

those with multiple authors. In fact, having articles with different number of authors in the corpus may produce more variety of moves and steps.

Hence, having considered section headings and type of research as the main selection criteria, 18 quantitative and 18 qualitative research reports were selected to construct a balanced corpus. These articles were first downloaded in pdf format, and then their Discussion and Conclusion sections were extracted and saved as text files. The distinctive features of quantitative and qualitative research expounded by Fraenkel and Wallen (2000) were taken into consideration in order to differentiate the former from the latter. In line with their definitions of both approaches to research, the quantitative studies selected for the corpus had quantitative methodologies with quasi-experimental, correlational, and pretest-posttest designs, and used instruments like structured questionnaires. In these studies, researchers adopted a deductive approach, stated hypotheses at the outset, controlled variables, carried out the analyses using statistical tools, and reported statistical summary of their results. In the qualitative studies selected for the corpus, on the other hand, researchers with a holistic approach were interested in finding their participants' beliefs, views, and concerns; they used interview transcripts, field notes and recordings, analyzed their data inductively, and used narrative descriptions rather than numbers to report what they observed.

The selected research articles came from six well-known journals, namely *Applied Linguistics, Language Learning, English for Specific Purposes, Modern Language Journal, TESOL Quarterly and System,* in the field of Applied Linguistics. These journals were listed among the top 25 % of journals indexed by Social Sciences Citation and considered that of category A in terms of their impact factors (Uysal, 2012). The impact factor reveals the citation percentage of journals and is accepted as an indication of the credibility of journals.

Table 2 shows the impact factors of the journals from which the research articles were selected for the corpus of this study:

Table 2. Impact Factor of the Journals from which the Research Articles were Selected for the Corpus

Journal	Impact Factor (for the year 2012)
Applied Linguistics	1.500
Language Learning	1.318
English for Specific Purposes	1.146
Modern Language Journal	1.114
TESOL Quarterly	0.969
System	0.692

Source: Thomson Reuters 2013, Journal Citation Reports

Each one of these journals was represented with 6 articles in the corpus. In order to avoid biases, the articles were selected either from the different journal volumes or from the different issues of the same volumes. In other words, no more than one article came from the same issue, and no more than two articles related to the same subject area were selected for the corpus. The following table (Table 3) displays the distribution of corpus data by the journals and the years of publication. For easy identification during the analyses, the concluding sections that come from quantitative reports were codified RA1-18 while those that come from qualitative reports were codified RA19-36 (as shown in the shaded cells of Table 3). The former included 1,172 sentences with 34,767 words while the latter included 1,010 sentences with 31,505 words within the corpus. In total, the corpus data analyzed included 2,182 sentences with 66,272 words.

Table 3. Distribution of the Corpus Data by the Journals and Publication Dates

	2007	2008	2009	2010	2011	2012	Total
AL	RA 19	RA20	RA21	RA22	RA 1	RA2	6
ESP		RA23	RA24 RA25	RA 26	RA27	RA28	6
LL	RA3	RA4	RA5	RA6 RA7	RA8		6
MLJ	RA29	RA30	RA9	RA10	RA11	RA12	6
SYSTEM	RA31	RA32	RA33		RA34	RA13 RA14	6
TQ	RA15 RA35	RA16		RA17	RA 18	RA36	6
Total	6	6	6	6	6	6	6

AL: Applied Linguistics   MlJ: Modern Language Journal   TQ: TESOL Quarterly
ESP: English for Specific Purposes   LL: Language Learning

3.3   Data Analysis

The study employed both qualitative and quantitative data analysis methods. This section describes in detail the procedure followed to design a move coding scheme and how each unit was coded based on this scheme in line with the conventions of move analysis.

3.3.1   Move Structure Analysis of the Corpus Data

In this study, a combination of corpus–based and genre-based approaches in text analysis advocated by some authors (Upton & Connor, 2001; Flowerdew, 2005; Flowerdew & Forest, 2009) was adopted for the move structure analysis of the corpus data in Swales' tradition. As suggested by Flowerdew (2005), the integration of these approaches counteract the arguments that criticize the corpus-based methodologies on the grounds that they only call for the bottom-up type of approach to genre analysis that investigates the lexico-grammatical patterning at sentence level. In other words, with a 'bottom-up' approach researchers distinguish moves on the basis of their linguistic features. A 'top-down' approach in text analysis, on the other hand, takes the macrostructure of text as the basis for analysis by first relying on intuitive interpretations of content to determine the moves. This approach is based on cognitive judgment rather than the linguistic criteria and thus, is more functional and qualitative in nature, and more in line with the theoretical definition of moves (Biber, Connor and Upton, 2007). Therefore, the qualitative investigation of move structures in this study was conducted with a functional top-down approach under the guidance of the analytical steps suggested by Biber et. al (2007) in Table 4.

Table 4. General steps often used to conduct a corpus-based move analysis

Step 1:   Determine rhetorical purposes of the genre

Step 2:   Determine rhetorical function of each text segment in its local context; identify the possible move types of the genre.

Step 3:   Group functional and/or semantic themes that are either in relative proximity to each other or often occur in similar locations in representative texts. These reflect the specific *steps* that can be used to realize a broader *move.*

Step 4:   Conduct pilot-coding to test and fine-tune definitions of move purposes.

Step 5:   Develop coding protocol with clear definitions and examples of *move types* and *steps.*

Step 6:   Code full set of texts, with inter-rater reliability check to confirm that there is clear understanding of move definitions and how *moves/steps* are realized in texts.

Step 7:   Add any additional *steps* and/or *moves* that are revealed in the full analysis.

Step 8:   Revise coding protocol to resolve any discrepancies revealed by the inter-rater reliability check or by newly 'discovered' *moves/steps,* and re-code problematic areas.

Step 9:   Conduct linguistic analysis of move features and/or other corpus-facilitated analyses.

Step 10:   Describe corpus of texts in terms of typical and alternate move structures and linguistic characteristics.

_____

Source: Biber, Connor, and Upton (2007:34)

3.3.1.1   *Initial Coding of the Corpus Using Yang and Allison's Model*

Although Biber et al. (2007) indicate that there is no certain way of doing a move

structure analysis, they suggest an analysis procedure that starts with having a 'big picture

understanding of the overall rhetorical purpose of the text in the genre' (2007:33) before

segmenting it into its functional categories. As commonly accepted in the literature on genre

analysis of research articles, the main rhetorical purpose of the texts that comprise the corpus

of the present study is the discussion of research findings in relation to those of previous studies and drawing implications for practicing teachers and future researchers.

Taking this overall purpose into consideration, the corpus data was prepared for the analysis: First, the *sentence* was taken as the unit of analysis for an in-depth identification of move-step structure, because a sentence-level analysis would provide more insight into authors' communicative intentions to realize moves. Thus, every sentence within the corpus was segmented and numbered (e.g. RA1_25 referring to the sentence 25 in RA1). Then, the move structure model proposed by Yang and Allison (2003) was used as the initial coding scheme to determine the rhetorical function of each sentence (See Appendix B). This model included 7 Moves realized by 10 steps (4 steps for M3, 3 steps for M6 and 3 steps for M7). As stated earlier in this paper, the move-step categories in their model were identified by the analysis of RA discussions and conclusions published in journals within the field of Applied Linguistics. Besides, among the journals from which the articles were selected for their corpus were *Applied Linguistics, English for Specific Purposes, and TESOL Quarterly* similar to the present study. Therefore, their model among others was selected to be used as a starting point for the investigation of the discussion moves in the present study.

During the initial coding of the corpus, some sentences could be assigned to neither of the categories identified in Yang and Allison's model mainly due to the lack of a relevant category that defined the observed functions. These sentences were marked to be reevaluated later with a second rater. Once the initial coding of the whole corpus was completed, a second rater holding a PhD degree in Applied Linguistics and having conducted corpus-based research in academic writing was asked to code the data to assess the coding reliability at sentence-level analysis. This rater who had been well-informed of the purpose of the study first examined the move-step definitions and examples of Yang and Allison's model and

coded a couple of texts together with the researcher to get training on move-step analysis. Then, he pilot-coded one of the texts in the RA corpus independently to calculate *Cohen's kappa (k)* that is a chance-corrected measure of inter-rater reliability as will be explained later in this paper. This first attempt to calculate the degree of coding agreement between the two raters on this particular text revealed an inter-rater reliability of .41, which showed an agreement on the 20 of the 38 sentences found in the text. This low agreement was observed to have been caused by the difficulty of determining the functions of some sentences that could be assigned to more than one move category depending on how they were viewed by the coder: as individual sentences or within the context of immediate sentences. Hence, in an attempt to set consensual coding criteria, the following decisions were made regarding the cases with multiple functions to reduce the coding challenges and increase the reliability:

1. In cases where a sentence seemed to have more than one move, this sentence would be assigned to the most salient move considering the purpose of the author(s) in uttering that statement (Holmes, 1997; Yang and Allison, 2003).

2. In cases where a sentence had a linking word that built connection with the preceding and following sentences, then the move category of the immediate context of this sentence would be considered to determine its move-step category.

3. In cases where the dependent and independent clauses of a sentence had different moves, the sentence would be assigned to the move category of the independent clause; but that of dependent clause would also be hand-tagged to reveal the move combinations that might occur.

4. In cases where sentences separated with a comma (,) and/or conjunctions (e.g. *and, but* etc.) had different moves, they would be considered to have more than one move.

5. The sentences separated with punctuation marks like colon (:), semi-colon (;), and an underscore (_) would be considered as separate sentences in cases where they could stand alone as full sentences with a move.

6. The sentences within quotations would not be considered for move analysis.

Taking these cases into consideration, all of the texts in the corpus were coded by the second rater, followed by after-coding discussion sessions with the researcher. During these labor-intensive sessions, which were occasionally held on Skype, raters could reconcile the coding differences to some extent through discussions that followed multiple readings of the sentences.

However, a considerable number of sentences could not be assigned to the available categories of Yang and Allison's Model. For example, according to their model, the authors referred to literature only for the purpose of comparing their results when *commenting* on them *(Move 4, Step 2)*. However, in the corpus data it was observed that the authors often referred to literature for multiple purposes: to remind the topic related discussions in the literature as background information; to seek support for the claims they make and the implications they draw; to interpret their results with other authors' views or the findings of other studies; to account for their results with other authors' explanations for similar results; and to justify the methodology they used by referring to other studies. Besides, in some rare cases, the authors of the RAs made comments on the results of other studies in the literature. They even evaluated the explanations offered by other authors. These observations, hence, led to the redefinition of the available move-step categories and formulation of new steps and substeps to accommodate authors' communicative intentions identified in the corpus data.

3.3.1.2 Constructing the New Model with Redefined Move-Step Categories

After the examination of the functional themes of sentences that could not be classified by the 14 rhetorical categories of Yang and Allison's Model, new steps and substeps were formulated and whole corpus was recoded considering the emerging new categories. In other words, it became the new coding scheme for the move analysis. During this recoding process, the definitions of the newly suggested categories were constantly fine-tuned to encapsulate the function of each sentence within the corpus. Once the coding was completed, the final version of the new model included 25 rhetorical categories realized by 6 Moves, 17 steps and 14 substeps (Appendix C). Since the new coding scheme has emerged as a result of the move analysis, it is presented in the results section of this dissertation (see Section 4.1) where each step and substep will be defined in detail and illustrated with corpus data.

The new model was, then, piloted on the Discussion and Conclusion sections of six RAs, three of which were published in *Studies in Second Language Acquisition* in 2012 and other three in *English for Specific Purposes* in 2013. Although these articles were not part of the corpus of the study and not considered for the lexico-grammatical analyses, this piloting was needed to see the extent to which the new move-step categories driven from a certain corpus could identify the rhetorical structure of other papers published in different journals. With the help of piloting, the scheme then in progress was given its final form in a way that it included all the possible steps that the authors used to realize moves in the discussion with conclusion sections of their RAs.

Finally, the corpus recoded using the new model was once again reviewed by the second rater during the "face-to-face" sessions in which he interpreted the rhetorical purpose of each unit to confirm the assigned move-step categories or to suggest changes in them. In cases where any modifications were suggested by the second rater, coded data were revised

over and over to apply the changes. Therefore, the construction of the new model could not be finalized until a joint decision was made regarding the move-step category of each sentence and their definitions. These sessions with the second rater, both during the initial coding of all data based on Yang and Allison's Model and the second coding based on the new model with redefined categories, took approximately 200 hours of discussion (for the curious reader), and formed the most time-taking, and challenging part of the study.

At this stage of coding, the agreement between the coders was checked periodically as described in the following section.

3.3.1.3 *Inter-rater Reliability Checks*

Although the second rater performed coding in the presence of the researcher in order to share his judgments immediately and to make a joint decision by resolving any dispute through discussion, a couple of texts in the corpus were randomly selected for independent coding to make periodic inter-rater reliability checks. For this purpose, Cohen's Kappa statistic, which is one of the most commonly used methods of computing agreement estimates of inter-rater reliability (Stemler, 2004), was used to compute the degree of agreement between the raters. Unlike the simple calculations of percent-agreement between the raters, Cohen's Kappa statistic accounts for the possibility that raters may agree with each other a certain percentage of the time on the basis of chance alone. Therefore, its coefficient reveals the proportion of joint judgments after chance agreement is eliminated from consideration (Cohen, 1960). The kappa coefficient between 0.21-0.40 is commonly interpreted as showing *fair* agreement, the values between 0.41-0.60 *as moderate* agreement, values between 0.61- 0.80 as *substantial,* and those over 0.81 as *almost perfect* agreement between the raters (Stemler, 2001; Lim, 2010).

For inter-rater reliability check, the second rater coded two texts, one representing the quantitative report and one representing the qualitative report, independently in line with the assumptions of Cohen's Kappa statistic. For the quantitative text which included 74 sentences and the qualitative text that included 87 sentences, the reliability coefficients calculated at step level considering the 25 categories of the coding scheme (see section 4.1.1) were found to be .59 and .54, respectively. These values are commonly interpreted as revealing a *moderate* agreement between the raters (Stemler, 2001). The reliability co-efficient calculated for the same texts at move (6 categories) level, however, revealed a better agreement between the raters with a value of .70 for the quantitative and .74 for the qualitative text. This difference between the reliability coefficients at step level and move level shows that the raters who have better agreement at move level may classify sentences as belonging to different steps of the same move. Since this reliability check was carried out at the early stages of coding data based on the new scheme, disagreement at the step level was used to refine the definitions of steps and substeps.

When the inter-rater reliability was computed for another independently coded qualitative text a couple of weeks later, the degree of agreement between the coders was found to be .81 (Appendix D shows K calculation for this text, RA24, using the formula $K = \frac{f_0 - fc}{N - fc}$ where *N* is the total number of sentences in the text coded independently; f_0 is the sum of sentences on the move-step categories on which both raters agreed in this text; and f_c is the sum of expected agreement by chance). Thus, this value reflected the observation that as the coders spent more time discussing the rhetorical purpose of each sentence, consistency in independent decisions increased. As a matter of fact, the inter-rater reliability checks were not conducted to share the load of coding work with another rater; but to check the clarity of move-step definitions.

The inter-rater reliability was also checked with a third rater who had been an experienced teacher and a graduate student enrolled in a MA program in Applied Linguistics. He was asked to code a couple of texts after being given training in move analysis with the new coding scheme. During the first meeting with him, he was given detailed information regarding how move structure analysis was conducted to identify the function of each segment within this genre. Then, he was invited to the discussion sessions where he observed how coders interpreted the communicative function of each sentence to determine its move-step category. After being involved in a few hours of discussion sessions through Skype as an observer, he was asked to indicate whether he would agree with the raters on their judgments or not. Then, he was finally asked to code two randomly selected texts, one quantitative-one qualitative, independently to compute the inter-rater reliability. The calculations revealed a reliability of .71 for the quantitative text and .70 for the qualitative text at step level. During the after-coding discussions, the third rater admitted that he miscoded or misinterpreted the functions of some sentences and expressed his agreement with the researcher on the classification of those sentences. That experience showed the importance and necessity of having discussions after coding process and is suggested as a research implication of this study for move analysis.

The main steps of the coding process described above are summarized in Figure 1:

Figure 1. The summary of the coding process that led to the development of new coding scheme

Initial coding of all data based on Yang & Allison's Model by the researcher (Independent Coding)

↓

Training of the second coder & pilot coding by the second rater

↓

First attempt to calculate inter-rater reliability

↓

Re-setting coding criteria

↓

Coding of all data based on Yang & Allison's Model by the second coder (Independent Coding of each RA followed by immediate discussions)

↓

Review of the functions of unclassified sentences

↓

Defining new emerging categories

↓

Coding of all data based on the new model by the researcher (Independent Coding)

↓

Coding of all data based on the new model by the second coder (Independent Coding of each RA followed by immediate discussions + Inter-rater reliability checks)

↓

The New Model is developed as the final version of coding scheme to be used in this study

3.3.1.4 *After-Coding Review of Corpus Data*

Once the coding was completed with consensual agreement on authors'
communicative intentions in each sentence, all of the coded sentences were listed under their
move-step categories. Thus, sentences identified in each RA as having *move 1*, for instance,
were put together for a final review of their functions. Although sentences were
decontextualized this way, it enabled the detection of those that did not seem to share the
same function as the other sentences with identical moves. After checking its context in the
article again, such sentences were, then, sent to the appropriate move-step categories if
miscategorized.

The listing of each sentence by their functional categories also enabled us to observe if
a particular step/substep occurred in enough number of RAs to be considered conventional in
Discussion sections. At this point, occurrence in 60 % of the corpus, as suggested by
Kanoksilapatham (2005), was taken as a cut-off point. In other words, the individual steps
and substeps that occurred in at least 60 % of 18 RAs within each subcorpus were considered
'conventional' to use as they reflect the community's conventional discourse norms. Whereas
those that fell below 60% were considered 'optional'. Optional moves and steps can be seen
as alternative functional categories that are available to the authors to discuss their findings.
The authors are not necessarily obliged to use them to get acceptance from the discourse
community. The moves and steps identified in all (100 %) of the RAs, however, are
considered 'obligatory', i.e. their use is expected by the discourse community, because they
are the sine qua non of the discussion sections. The obligatory and conventional moves and
steps encapsulate the communicative functions of the genre in which they were identified
more than the optional ones. (See Appendix E for move-step categories observed at least
once in each RAs).

3.3.1.5 *Searching for Move Sequences*

Move sequences were investigated at three levels (move, step and substep) through AntConc 3.4.1w, a freeware corpus analysis tool developed by Laurence (2011). All RAs were rearranged to include the strings of codes without sentences and screened through the *N-gram* facility of AntConc which was basically used to find lexical bundles in a corpus. *N* in n-gram usually varies from 2 to 5 referring to the sequences of word-pairs (2-grams), word triplets (3-grams) and so on. Thus, the most frequent strings of 2-to- 5 moves, steps or substeps that follow each other in the corpus data were searched for. As the corpus of the study is relatively small, those that co-occur at least 5 times in a range of 3 RAs within each sub corpora were tabulated to see the move categories that co-occur.

3.3.2 Quantitative Analysis of Data

The occurrences of each rhetorical move were analyzed at step and substep level in relation to the type of research (i.e. quantitative versus qualitative) with reference to their percentage of occurrences (See Appendix F and G). In order to see if there is a statistically significant difference between the quantitative and qualitative RAs in terms of the occurrences of discussion moves, a non-parametric independent samples t-test was needed because the Shapiro-Wilk Test of Normality revealed a violation of the normality assumptions for all distributions ($p<.05$). In other words, the percentages of occurrences were not normally distributed in this study; therefore, a chi square test was run using SPSS 21.0 software. Where the expected count of move-step occurrences is less than 5, Fisher's Exact Test was used.

3.3.3 Identification of the lexico-grammatical features of Moves

The lexico-grammatical analysis of the study was limited to the investigation of the formulaic nature of RA discussions. The study aimed to find recurrent grammatical structures associated with six moves identified in the first part of the study. As the corpus data had already been analyzed and classified by their functional categories through move analysis, a function-first approach advocated by authors like Wray (2002) and Durrant & Mathews-Aydınlı (2011), which was mentioned earlier, was adopted naturally. Due to the small size of corpus data that represent each one of the 25 functional categories, the analysis was conducted at move level to identify high-frequency formulaic bundles based on their grammatical structures. Similarly due to the same reason, the study did not aim to compare the use of formulaic bundles in quantitative and qualitative RAs (see Appendix L for move sizes). Each word including punctuation in the corpus data, was first tagged by their part-of-speech (PoS) categories, through Stanford tagger (Toutanova, Klein, Manning, and Singer, 2003). The reason for tagging punctuation lies in its contribution to the establishment of meaning in written discourse. Besides, it reveals where in a discourse a certain bundle occurs as pointed out by Qin (2014) who also considered punctuation as part of formulaic bundles in academic writing.

A *Key-bundle Analyzer Program* developed by Karabacak (2009) was, then, used to identify the key PoS categories that are statistically typical of the sentences in each move with reference to those in other moves. This program used Ted Dunning's log-likelihood formula explained in Karabacak's 2009 study to calculate a statistical value that determined the keyness *of PoS n-tags* in each move. At this point, the cut-off log-likelihood value to determine the *keyness* of PoS categories in certain moves was decided to be 6 (including 5.5) or higher (Karabacak, 2009). *PoS n-tag* here refers to *n* number of co-occurring tagged categories. In this study, 5-PoS and 4-PoS tag categories that yield 5-word and 4-word

combinations were focused, because they already have many 3-word lexical and grammatical bundles in their context, and they present enough number of units to analyze in terms of structure and function in relation to that of the specific move.

On the other hand, a minimum of 5 times of occurrence in at least 3 different articles was determined to be the cut-off value for frequency of occurrence to be able to confirm the keyness of a sequence in a certain move. As the frequency threshold suggested by Biber, Johansson, Leech, Conrad, and Finegan (1999) is at least 10 times of occurrence in five different texts in a million-word corpus, the normalized value for the 66,272-word corpus of the study is almost equal to one time of occurrence. Therefore, setting the cut-off value at 5 times of occurrence is quite a conservative approach that will yield more reliable results.

The PoS tag sequences (tag-bundles) found to be typical of a certain move by these cut-off values were later searches to extract the real word sequences from the corpus through the use of AntConc 3.4.1.w.

CHAPTER IV

RESULTS AND DISCUSSION

In this study, the first level of analyses was conducted to identify the genre-specific rhetorical structure of the concluding sections of RAs published in the field of ELT/Applied Linguistics. Based on this analysis, a new move structure model was proposed as the new coding scheme to encapsulate the communicative purposes that writers meant to achieve while discussing their results. In this regard, the study aimed to answer the following research questions:

1. What are the genre-specific rhetorical features of the discussion with conclusion sections of the published research articles in the field of ELT/ Applied Linguistics?

1.1. What types of moves are there in the discussion sections of the published research articles in the field of ELT/ Applied Linguistics? Which steps and substeps do the authors use to realize these moves?

1.2. What are the obligatory, conventional, and optional moves identified in the published RA discussions in the field of ELT/ Applied Linguistics?

1.3. Are there any differences between the quantitative and qualitative RA discussions in terms of move frequency? If so, what are they?

1.4. Are there any differences between the quantitative and qualitative RA discussions in terms of the move-step sequences? If so, what are they?

The second level of analyses was conducted using corpus linguistics methodologies to investigate the formulaicity in moves identified in the first part of the study. With this respect, the following research questions were addressed:

2. What are the lexico-grammatical features of the moves identified in the discussion with conclusion sections of the published research articles in the field of ELT/ Applied Linguistics?

2.1. Are there any strings of 5-PoS tag bundles that are typical of each discussion Moves? If yes, what are these grammatical and lexical structures?

2.2. Are there any strings of 4-PoS tag bundles that are typical of each discussion moves? If yes, what are these grammatical and lexical structures?

4.1 Results of Move Structure Analysis

This section reports and discusses the results of the move structure analysis to answer the research questions, 1.1, 1.2, 1.3 and 1.4.

4.1.1 Moves and Steps Identified in the Study: The categories of the New Coding Scheme

Move analysis of the discussion with conclusion sections of the RAs published in the field of ELT/Applied Linguistics identified 6 rhetorical moves (M1, M2, M3, M4, M5, M6) realized by 17 steps (M1S1, M1S2, M1S3, M2S1, M2S2, M2S3, M3S1,M3S2, M3S3, M3S4, M4S1, M4S2, M5S1, M5S2, M5S3, M6S1, M6S2) and 14 substeps (M1S2A, M1S2B, M3S1A, M3S1B, M3S2A, M3S2B, M3S2C, M3S3A, M3S3B, M3S3C, M6S1A, M6S1B, M6S2A, M6S2B). These moves, steps and substeps formed the new coding scheme with 25 categories to classify the communicative functions of the 2,182 sentences within the corpus. In the 1,172-sentence subcorpus of the quantitative RAs and 1,010-sentence subcorpus of the qualitative RAs, 1,222 and 1,061 moves were observed, respectively. 50 sentences in each subcorpus were classified as contributing to two moves while 1 sentence in the qualitative data included three moves. Thus, move analysis revealed that the authors of the 36 RAs

selected for the corpus of this study used a total of 2,283 moves to discuss their results and to conclude their studies.

Before defining each category of the new coding scheme, the number and percentage of the quantitative (N=18) and qualitative (N=18) RAs in which they occurred are shown in Table 5 (These numbers come from the table given in Appendix E). The Table 5 also shows their rank and whether they are obligatory (if observed in 100 % of RAs), conventional (in more than 60% of RAs), or optional (in less than 60% of RAs) in each subcorpus.

Table 5. The Move-Step Categories Identified in the Corpus

	QUAN RAs			QUAL RAs		
	F Percentage (N)		*Rank*	*F Percentage (N)*		*Rank*
M1S1	100 (N=18)	**Obligatory**	**1**	55.56 (N=10)	Optional	6
M1S2A	83.33 (N= 15)	Conventional	4	83.33 (N=15)	Conventional	**3**
M1S2B	44. 44 (N= 8)	Optional	9	33.33 (N= 6)	Optional	10
M1S3	27.78 (N= 5)	Optional	12	50 (N= 9)	Optional	7
M2S1	33.33 (N= 6)	Optional	11	0 (N= 0)	Not observed	14
M2S2	94.44 (N= 17)	Conventional	**2**	77.78 (N=14)	Conventional	4
M2S3	50 (N= 9)	Optional	8	27.78 (N=5)	Optional	11
M3S1A	100 (N= 18)	**Obligatory**	**1**	94.44 (N=17)	Conventional	**1**
M3S1B	66.67 (N= 12)	Conventional	6	55.56 (N= 10)	Optional	6
M3S2A	77.78 (N= 14)	Conventional	5	55.56 (N= 10)	Optional	6
M3S2B	50 (N= 9)	Optional	8	27.78 (N= 5)	Optional	11
M3S2C	16.67 (N= 3)	Optional	13	11.11 (N= 2)	Optional	12
M3S3A	100 (N= 18)	**Obligatory**	**1**	66.67 (N=12)	Conventional	5
M3S3B	61.11 (N= 11)	Conventional	7	38.89 (N= 7)	Optional	9
M3S3C	38.89 (N= 7)	Optional	10	11.11 (N= 2)	Optional	13

M3S4	83.33 (N= 15)	Conventional	4	55.56 (N= 10	Optional	6
M4S1	50 (N= 9)	Optional	8	44.44 (N= 8)	Optional	8
M4S2	77.78 (N= 14)	Conventional	5	83.33 (N=15)	Conventional	**3**
M5S1	83.33 (N= 15)	Conventional	4	44.44 (N= 8)	Optional	8
M5S2	50 (N= 9)	Optional	8	38.89 (N= 7)	Optional	9
M5S3	38.,89 (N= 7)	Optional	10	11.11 (N= 2)	Optional	12
M6S1A	94.44 (N= 17)	Conventional	**2**	83.33 (N=15)	Conventional	3
M6S1B	100 (N= 18)	**Obligatory**	1	50 (N= 9)	Optional	7
M6S2A	88.89 (N= 16)	Conventional	**3**	94.44 (N=17)	Conventional	**1**
M6S2B	61.11 (N= 11)	Conventional	7	88.89 (N=16)	Conventional	**2**

As the Table 5 shows, the study identified 4 obligatory (M1S1, M3S1A, M3S3A and M6S1B), 11 conventional (M1S2A, M2S2, M3S1B, M3S2A, M3S3B, M3S4, M4S2, M5S1, M6S1A, M6S2A and M6S2B), and 10 optional categories (M1S2B, M1S3, M2S1, M2S3, M3S2B, M3S2C, M3S3C, M4S1, M5S2, and M5S3) for the quantitative RAs. On the other hand, the study identified no obligatory categories, but 8 conventional (M1S2A, M2S2, M3S1A, M3S3A, M4S2, M6S1A, M6S2A, and M6S2B) and 16 optional categories (M1S1, M1S2B, M1S3, M2S3, M3S1B, M3S2A, M3S2B, M3S2C, M3S3B, M3S3C, M3S4, M4S1, M5S1, M5S2, M5S3, and M6S1B) for the qualitative RAs. The second step of M2, however, was not observed at all in the qualitative RAs. These variations in relation to the quantitative and qualitative approach to research will be discussed below after each rhetorical category is defined and illustrated with sentences from the corpus data:

Move 1 (M1) – Setting Background

This move was realized by three steps that aimed to provide different types of background information about the study:

<u>Move 1, Step 1 (M1S1) – Methodological background</u>

M1S1-Providing methodological information regarding the study is similar to *Background Information Move* identified in RA discussions in the previous studies (e.g., Dubley-Evans, 1994; Peacock, 2002; Yang and Allison 2013). The authors used M1S1 to restate methodological information that had already been mentioned in the previous sections of the RAs. This information included the aims and purposes of the study (excerpt 1), the hypotheses and research questions addressed by the study (excerpts 2 and 3), the context and participants of the study (excerpts 4 and 5) as well as the authors' initial expectations regarding the findings of the study (excerpt 6). The following excerpts (1-6) illustrate the use of M1S1 both in the quantitative (samples from RA1-18) and qualitative RAs (samples from RA19-36) with lexical signals bolded:

(1) ***The first goal of this study*** *was to assess the independent contribution of morphological awareness to reading comprehension when the variance due to other key reading - and language - related variables was controlled* (RA11_1)

(2) ***The first hypothesis predicted*** *a significant relationship between reading span and inferential comprehension accuracy but not literal understanding (RA18_2)*

(3) ***The third research question concerned*** *the extent to which any effect for written CF is mediated by language analytic ability (RA15_31)*

(4) *In our case, the L2 learners had spent, on average, five college semesters in courses that exposed them to input, grammatical terminology, and activities of the type administered in our study* (RA5_18)

(5) ***The participants in this study had*** *completed pedagogical coursework and used contemporary textbooks that subscribed to principles of communicative language teaching as well as standards developed by national FL organizations* (RA29_55)

(6) ***We began this study with*** *relatively simplistic* ***predictions of*** *what we would find: Disciplines in the humanities* ***would be*** *the most language dependent, while the sciences and math* ***would be*** *the least language dependent, with the social sciences being somewhere in between* (RA22_6&7)

As observed in Amnuai and Wannaruk (2013), M1S1 was the most frequently used opening move in the corpus data. Twelve (10 quantitative and 2 qualitative RAs) of the 36 RAs began with M1S1, which was frequently followed by the steps of *discussing* (M3) and *reporting the results* (M2) moves due to the authors' desire to remind readers the purpose and the methodology of the study before interpreting the findings. Besides, the authors also needed to point to some methodological details while evaluating their study (M5) especially in terms of the methods used, and while discussing their pedagogic and research implications (M6). The bold sentences in the following excerpt show how authors relate methodological information to the implications they deduct for EFL teachers and learners before concluding the paper:

(7) *Second, this study provides further evidence that EFL teachers and learners require support when transitioning from traditional L2 teaching methods to task-based language teaching, as previous studies have reported (e.g., Li, 1998, Mok-Cheung, 2001).* ***In this context, the teachers were provided with supplementary***

materials and an introductory workshop to raise their awareness about the

*principles of task-based language teaching. **Introductory units and supplementary***

materials were added to the course to help learners recognize the course content and

***teaching approach.** Integrating cognitive and metacognitive learning strategies into*

task-based courses may be particularly helpful for learners who are accustomed to

teacher-fronted grammar-based courses (RA35_8-11)

Move 1, Step 2 (M1S2) – Theoretical background

The other type of background information provided by the authors was **theoretical**. In order to set conceptual background for the study, the authors used **Move 1**, **Step 2 (M1S2)** by **A) describing the established knowledge and stating the gaps (M1S2A)** and **B) making claims and deductions based on the established knowledge (M1S2B)**. Unlike M1S1 that included information regarding the method of the study to be reported, the sentences with **M1S2** were observed to have functions that are more similar to those of statements uttered to establish research territory in the introduction sections of the RAs as identified by Swales (1990). Using **M1S2A** the authors often made general statements (topic generalizations) about the knowledge and practice to lay the basis for the interpretation of their findings. In other words, they provided theoretical background by referring to the topic-related discussions and the acknowledged claims of other authors in similar studies. Besides, the authors used M1S2A in order to state the gaps in the literature, which is another common move observed in the introduction sections to establish the niche. The following excerpts (8-13) taken from the corpus data illustrate these functions of M1S2A:

(8)   *Although **it has been widely argued that** approaches to second-language*

instruction should ensure that learners develop a rich repertoire of formulaic

*sequences (Ellis 2005:210–11), **there has been little empirical evidence to show that***

the formulaicity of learner language directly contributes to communicative competence (RA1_6)

(9) ***Previous research has shown that*** *learners' language proficiency plays an important role in their ability to benefit from feedback* (RA3_C92)

(10) ***Studies have not made a distinction between*** *"time on task"* ***and*** *engagement and cognitive processing time* (RA12_C59)

(11) ***A general consensus among researchers is that*** *probing and collaborating stances are more beneficial to student writers than prescriptive and authoritative ones (Lockhart & Ng, 1995a; Mangelsdorf & Schlumberger, 1992)* (RA23_39)

(12) ***Many studies of*** *implicit and explicit knowledge in the field of SLA* ***often fail to acknowledge*** *the different levels of consciousness and of verbalization ability* (RA30_36)

(13) ***In his description of*** *the characteristics of explicit knowledge of language, Ellis (2004, 2005)* ***stressed the fact that*** *this knowledge is potentially verbalizable* (RA30_41)

While setting theoretical background, the authors also made **generalizations and claims based on the established knowledge (M1S2B)** in relevant literature, not on the findings of their study yet as illustrated in the following statements (14-19):

(14) *We have reason to believe that a purely incidental learning situation would have led to even lower recall scores (RA2_C26)*

(15) *Although identification of individual morphemes must occur first for any type of morphological analysis to take place, it is through the semantic processing of morphemes (interpretation) that the access of word meaning takes place, which then ultimately aids text comprehension* (RA11_52)

(16) ***Considering relevant research findings, however, it is premature to claim that*** *interpretation and identification are two empirically distinct constructs* (RA11_53)

(17) *Making research public, as I argued earlier, is a defining characteristic of research (*RA21_10)

(18) *In the case of repetitions,* ***it can be argued that*** *the participants evaluate the previous formulation as they repeat it* (RA30_11)

(19) *Thus, as van Lier points out, linguistic and metalinguistic awareness are a natural result of literacy* (RA30_33)

The models suggested previously by Yang and Allison and Dudley-Evans recognized methodological and theoretical background as part of their *Background Information Move*. However, in the new model suggested here these two were defined as different steps of the same move, because providing methodological background (M1S1) involved reporting what is done by the author in the present study whereas providing theoretical background (M1S2) involved claims based on what was done by the other authors in previous studies. With this regard, the functions of M1S2 are equivalent to the functions of a move identified by Kanoksilapatham (2005) in RA discussions within the field of biochemistry. In his corpus, he defined authors' purposes to *contextualize the study (Move 12)* through the steps of describing *the established knowledge (S1)* and *presenting generalizations, claims, deductions or research gap (S2).* He observed authors' tendency to start using this move by first presenting

generalizations (S2) supported with some previous work (S1) and followed by the statements of gap in the literature to assert the centrality of the topic of their study. Similarly, the authors in the present corpus generally used the strategies of M1S2 together to contextualize the study by pinpointing its centrality before presenting what is suggested by its individual findings.

Move 1, Step 3 (M1S3) – Organizational background

Making statements about the organization of the information to be presented is another type of background information identified in the corpus data. These statements hold information regarding the rhetorical organization of what will be presented in the text. In other words, the authors use this step to announce what is next, and what will be displayed in tables/graphs and appendices as seen in the following examples (20-25):

(20)   *We now turn to the amount of "scaffolding" needed for the intervention (Research Question 2) (RA4_21)*

(21)   *In the next section we also consider the potential benefits of aural uses of written texts, based on observations of teacher practices in our corpus (RA9_43)*

(22)   *In order to illustrate the potential of keystroke logging for analysis of individual writers, we have summarized in Table 3 the automatic and manual analyses of three texts produced by one writer, Jonas (RA16_10)*

(23)   *In what follows we will interpret our students' (novice scientists') practices and beliefs by first acknowledging the novice ' s perspective, which then leads to a discussion of the formulaicity of scientific writing and originality in science (RA19_11)*

(24)   *I will therefore discuss each content category in turn (RA28_10)*

(25) ***The following section explores*** *how such a third space might be constructed in*
Hong Kong schools to narrow antagonistic identity relations (RA36_36)

<u>Obligatory, Conventional, and Optional M1 categories and their frequency of
occurrence in the quantitative and qualitative RAs</u>:

Given the functions of M1 steps, the number and the percentage of the quantitative and
qualitative RAs which included them and their frequency of occurrence in two subcorpora
have been displayed in Table 6. The percentages of each move in each RA, and the central
tendencies of each move based on these percentages, however, are shown in Appendix G and
Appendix H respectively. Table 6 also shows the p values of the chi-square analysis.

Table 6. Frequency of Occurrence of M1 at Step and Substep Level

	N (%) of RAs containing M1 Steps in the Corpus (N=36)	QUAN RAs (N=18)			QUAL RAs (N=18)			
		N (%) of RAs containing M1 steps	f in QUAN RAs	% of all QUAN Moves	N (%) of RAs containing M1 steps	f in QUAL RAs	% of all QUAL Moves	X^2
M1S1	28 (78) **	18 (100)*	78	6.38	10 (56)	37	3.48	.007 (df=25)
M1S2 (M1S2A & M1S2B)	32 (89) **	17 (94) **	107	8.75	15 (83) **	132	12.44	-
M1S2A	30 (83) **	15 (83) **	88	7.20	15 (83) **	96	9.05	.576 (df=29)
M1S2B	14 (39)	8 (44)	19	1.55	6 (33)	36	3.39	.733 (df=14)
M1S3	14 (39)	5 (28)	9	0.74	9 (50)	12	1.13	327 (df=13)

*obligatory, ** conventional*

Table 6 shows that all of the quantitative RAs included **M1S1** (Also see Appendix E). However, it occurred only in 56% of the qualitative RAs, which is below the cut-off point of 60%. That means **M1S1- Reviewing methodological information regarding the study** is an *obligatory* move that needs to be used in the quantitative RAs, but *optional* in the qualitative RAs published in the field of Applied Linguistics/ELT. In the meantime, the values obtained through the frequency counts reveal that M1S1 was observed 115 times in the corpus data. Seventy eight of 115 occurrences were identified in the quantitative subcorpus whereas 37 were in the qualitative subcorpus (See Appendix F for the distribution of these occurrences by RAs).

The table also shows that the authors provided **theoretical background information (M1S2)** through either one of its substeps in 94 % of the quantitative and 83 % of the qualitative RAs (see the second row in Table 6). In other words, 17 of the quantitative and 15

of the qualitative RAs included at least one substep of M1S2. It was *conventionally* realized **by referring to established knowledge** (**M1S2A)** in the 83 % of both subcorpora. Although M1S2A was observed in the same number (N=15) of the quantitative and qualitative RAs, it was identified 88 times in the former, and 96 times in the latter. In fact, among all steps and substeps of M1, **M1S2A** was the most frequently observed M1 step in both subcorpora. On the other hand, the other substep, **M1S2B – making claims based on established knowledge** appeared in 14 (8 quantitative and 6 qualitative) of the 36 RAs as an *optional* step of M1 with 44% and 33% of occurrences in the quantitative and qualitative subcorpus, respectively. Although M1S2B was observed in fewer numbers of the qualitative RAs (6 RAs vs.8 RAs), it was used more often by the authors of the qualitative RAs (36 times vs. 19 times). Thus, the qualitative RAs were found to include more instances of providing theoretical background for the study.

In order to see if these differences are statistically significant, the percentages of the total number of quantitative and qualitative moves that M1 steps made up were calculated, because the quantitative and qualitative RAs of the study did not have equal number of sentences. These calculations are shown in Figure 2 (see Appendix H for the descriptive statistics).

Figure 2. Percentage of quan and qual moves that are made up of M1 Steps

81
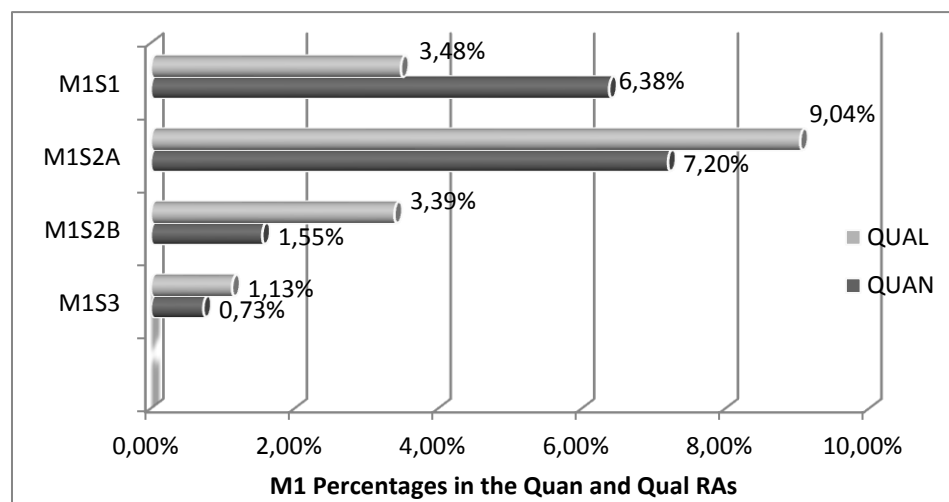
As Figure 2 shows, 78 M1S1occurrences made up 6.38% of 1222 moves (M= 6.76, SD=4.219) observed in the quantitative data while 37 M1S1 occurrences made up 3.48% of 1061 moves (M= 3.34, SD= 4.32) observed in the qualitative data. These numbers of occurrences showed that the authors of the quantitative RAs provided more methodological background information than the authors of the qualitative RAs. As the Fisher's exact *p* value for the quantitative–qualitative comparison of M1S1 occurrence is .007,

(See Table 6 & also Appendix I), this difference is significant, *(25, N=36)=32.00, p< .05*, showing the relation between authors' tendency to give methodological background and the type of research.

This tendency can be attributed to the underlying assumptions that determine the research purpose of quantitative and qualitative studies. As opposed to qualitative research that is concerned with constructing *multiple realities* based on a naturalistic-phenomenological philosophy, quantitative research aims to reach an accurate *objective reality* based on a logical positivist philosophy (McMillan & Schumacher, 1989). In order to achieve this aim, quantitative methods provide researchers with more established sets of procedures through which they can do experiments in highly controlled situations, and conduct reliability and validity measures to obtain findings that they can generalize from sample to the population. The methodological procedure followed, then, becomes important for authors to justify the objectivity and generalizability of their findings. Therefore, it is not surprising to identify the methodological background move in the discussion sections of all quantitative RAs with more frequency than in qualitative RAs.

On the other hand, the results regarding the use of M1S2 showed that the authors provided more theoretical background in qualitative RAs. As shown in the Figure 2, it accounted for 12% (9% by M1S2A and 3% by M1S2B) of all qualitative moves and 8.5% (7% by M1S2A and 1.5% by M1S2B) of all quantitative moves (See Appendix H for

descriptive statistics). Although this was not a statistically significant ($p> .05$) difference, its more frequent use in qualitative RAs can be attributed to the inductive nature of this approach to research in which authors may need to present more theoretical information while constructing a grounded theory. As a matter of fact, M1S2 was found to be the most frequently used step of M1 in both subcorpora. One possible explanation for this finding is that despite the importance of methodological background as discussed above, the purpose of M1S2 through which the authors provide theoretical background, contributes more to the overall rhetorical purpose of the Discussion sections. That is to say, the main purpose of the Discussion sections is not to describe the methodology but to interpret the findings in relation to the underlying theoretical knowledge. Therefore, M1S2 enables authors to discuss their findings within that framework, which in turn contributes to the contextualization of the study. This is in fact how authors situate their studies in the interest of discourse community as pointed by Kanoksilapatham (2005) and to the benefit of readers, because thanks to that theoretical framework, the findings make more sense to them.

As for the final step of M1, **M1S3**- rhetorical information -was identified 12 times in 50% of the qualitative RAs and 9 times in 28% of the quantitative RAs as an optional move. These numbers of occurrences made up 1.13% (M= 1.08, SD= 1.20) and 0.74% (M=.73, SD=1.39) of these subcorpora respectively and showed no statistical difference.

Move 2 (M2) – Reporting Selected Results

This move aimed to restate the crucial findings of the study through three steps:

Move 2, Step 1 (M2S1) – Reporting if the hypotheses are supported

Using **M2S1**, the authors indicated **whether or not results supported the hypotheses of the study.** The following examples (26-27) illustrate the sentences having this function in the corpus:

(26) ***The overall results confirm*** *one of our hypotheses* (RA13_7)

(27) ***The hypothesis was supported, in that*** *while both high - and low - span readers' performance in literal understanding was similar irrespective of the content type used, as shown by the insignificant difference between their means, high - span readers outperformed low-span readers in inferential comprehension (*RA18_3)

Move 2, Step 2 (M2S2) – Selected results with statistics

M2S2 aimed to restate the findings of the study **by providing relevant statistics, figures, and examples**. These were some of the significant findings that had already been reported in the Results section of the RAs; however, they were briefly repeated by the authors in the Discussion section especially within the context of Move 3. For instance, the sentences with M2S2 included the significant results of statistical analyses in quantitative articles; and what was observed by the author or stated by the participants of the study in qualitative articles. Such sentences did not involve any interpretation yet as seen in the samples given below (28-33).

(28) *In the present study, the rate of recasts that pushed the learners to modify their output was 72% and the rate of recasts minus prompts was 28%* (RA3_21)

(29) *The percentage of students who opted to continue with French beyond their AS exam was as follows: HSG: 69%; LSG, 28.2%; CG, 71. 8%* (RA4_27)

(30) *The results of the LCP **indicated** that the students participated in these activities for an average of 60.68 hours per week (M = 60.68, SD = 24.98)* (RA10_9)

(31) *At the same time, though, over 40 per cent of the teachers doing research said it was part of a course they were studying on (RA21_35)*

(32) *Hou, Burnerman, and Julius saw their fields of research as largely language - independent, and as far as their academic performance was concerned, they did not feel any setback because of their English proficiency* (RA22_20)

(33) *Indeed, all six Korean graduate students in my study identified both factors as influencing their own verbal participation in class discussions* (RA25_5)

Move 2, Step 3 (M2S3) – Integrated results

The final step of M2, **M2S3,** on the other hand, was used to **give integrated results**. The authors either listed a set of related results or provided an overall summary of such results as seen in the following sentences (34-37):

(34) ***Consequently***, *students with more autonomous feelings during English lessons were more self-determined **(RA14_15)***

(35) ***To summarize*** *the congruency effect, native speakers of English showed no difference in reaction time and error rate when they responded to English collocations that were congruent and incongruent to their counterparts in Japanese* (RA17_1)

(36) ***Overall***, *the responses from the heterogeneous sample of 505 teachers of English studied here indicated that their conceptions of research are aligned with conventional scientific notions of enquiry* (RA21_2)

(37) *Key ideas which resonated with teachers' notions of research were statistics,*
objectivity, hypotheses, large samples, and variables (RA21_3)

Obligatory, Conventional, and Optional M2 categories and their frequency of
occurrence in the quantitative and qualitative RAs:

Among the Move 2 steps defined above, **M2S2** through which the numerical findings
are reported with no interpretations yet corresponds to the *Reporting Results* move (M2) in
Yang and Allison's Model (See Appendix B). As seen in Table 7, **M2S2** was observed 135
times in the 94% (N=17) of the quantitative RAs and 143 times in the 78% (N=14) of the
qualitative RAs. With these values, M2S2 was the only *conventional* M2 step identified in
the study. These occurrences of M2S2 comprised 13.29% of all moves in the qualitative data
(M=11.50, SD= 12.51) and 10.96% of all moves in the quantitative data (M= 10.07,
SD=7.029) as shown in Table 7 and Figure 3.

Table 7. Frequency of Occurrence of M2 Steps

	N (%) of RAs containing M2 Steps in the Corpus (N=36)	QUAN RAs (N=18)			QUAL RAs (N=18)			X^2
		N (%) of RAs containing M2 steps	f in QUAN RAs	% of all QUAN Moves	N (%) of RAs containing M2 steps	f in QUAL RAs	% of all QUAL Moves	
M2S1	6 (33.33)	6 (33.33)	12	0.98	0	0	0	.303 (df=6)
M2S2	31 (86.11)[**]	17 (94.44)[**]	135	10.96	14 (77.78)[**]	143	13.29	.338 (df=31)
M2S3	14 (38.89)	9 (50)	18	1.47	5 (27.78)	7	0.66	.176 (df=17)

*obligatory, ** conventional

Figure 3. Percentage of quan and qual moves that are made up of M2 Steps
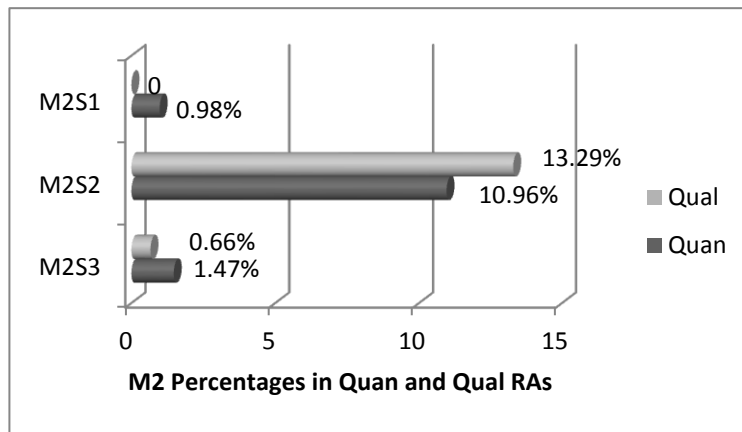
As for the other steps, M2S1 and M2S3 are not defined as different steps of M2 in Yang and Allison's Model. For example, statements regarding the hypotheses of the study being supported or not by the findings are not classified as part of a separate step of M2. However, in this study, such statements were observed only in the quantitative RAs and therefore, were classified by a different step of reporting results that is only available to the authors of the quantitative studies. However, as shown in Table 7 and Figure 3, 12 times of occurrence in 33% (N=6) of the quantitative RAs made up only 1% (M=1.37, SD=2.243) of all moves observed in this subcorpus.

The reason for the identification of M2S1only in quantitative RAs has to do with the deductive nature of quantitative studies as opposed to the inductive nature of qualitative studies. In quantitative approach, the researchers state the research questions and hypotheses deduced from theory at the onset of the study, make methodological decisions, and then conduct the research to test these hypotheses. The aim of the study is, in a way, to see if these hypotheses are confirmed, rejected or needed to be modified. Therefore, one can expect to find such statements in quantitative studies. In qualitative research, however, the researchers cannot start the study by stating the hypothesis first; instead they reach the theory by examining the phenomenon and suggest the hypothesis at the conclusion.

Finally, the findings regarding the use of M2S3 reveal that authors gave integrated results in 14 RAs (39%), 9 of which represented 50% of the quantitative subcorpus and 5 represented 28% of the qualitative subcorpus. Its 18 occurrences in quantitative RAs made up 1.47% (M=1.53, SD=2.14) of all quantitative moves and 7 occurrences in qualitative RAs made up 0.66% (M=.60, SD=1.22) of all qualitative moves (Figure 3).

M2S3 corresponds to Move 3 (*Summarizing results*) identified in Yang and Allison's Model. In this study, it was classified as part of *reporting result*s move, not as a separate move itself, because giving integrated results was observed as one rare way of reporting results, not as one of the major moves in this section. As a matter of fact, all of the results reported in Discussion sections are already stated in the Results section of the RAs, but since they are the salient ones that will be discussed by the authors, they are restated before the interpretation. Therefore, Move 2 was usually followed by the steps of Move 3 in the corpus of the study as will be discussed in section 4.1.2.

Move 3 (M3) – Discussing Results

As the main function of the RA discussions has been the discussion of the findings of the study, this move had naturally been the most frequently used one in the study. The authors of the RAs that the corpus of the study involved discussed their results by **interpreting** results, **explaining** the reasons for them, **comparing** them to those of previously reported studies, and **evaluating** them. These steps were achieved through the substeps described below:

Move 3, Step 1 (M3S1) – Interpreting results

Using this step, the authors interpreted their findings by **A) stating what the results suggested (M3S1A).** In other words, they explained what their results meant or how they

should be perceived within that field of research; they reached a conclusion based on their own interpretation and made general claims; and extended their arguments by raising some questions on which they also elaborated as shown in the following sentences (38-44).

(38) **There is thus evidence**, *albeit tentative,* **that** *strategy instruction with feedback that focuses on the link between strategy use and successful listening can have a positive impact on both listening performance and students' self - efficacy for listening, even with a relatively small amount of feedback* (RA4_34)

(39) *The long-lasting congruency effect on the ESL users ' error rate* **suggests that** *incongruent collocations are difficult to accept in the L2 mental lexicon, and acquiring this type of collocation takes a long time, requiring a massive amount of exposure to the L2* (RA17_4)

(40) **This indicates that** *L2 learners are initially dependent on the L1 mediation process, which resulted in the processing advantage of congruent collocations, but with the increase of exposure to and use of the L2, direct links between L2 collocations with concepts are formulated and L2 collocations come to be processed independently of the L1 lexicon* (RA17_9)

(41) *What about, for example, re-using someone else's description of background information in writing the introduction (Examples 3 and 4), or long stretches of wording when composing citations (our Examples 1 and 2), or experimental description in the methods section (Example 5), or the reporting of routinized results (Example 9), even if such re-use is not word-for-word copying but with some editing /adjustment?* (RA19_29)

(42-44) *Likewise, **it is possible to argue that** during revision, learners scan through their written text in order to find aspects that, in their view, can be improved. In other words, when learners make changes to their written text, they sometimes correct errors and, at the very least, they solve an aspect of their text that they perceive as problematic or improvable. Thus, reformulations to a written text are not only the result of noticing, which might only require implicit knowledge, but also of reflecting, which requires access to explicit knowledge* (RA30_66-68)

The authors also interpreted their results by **B) referring to established knowledge on topic (M3S1B).**   Unlike M3S1A, they benefited from other authors' views and the findings of other studies in literature in order to explain what their findings suggested; and therefore, they used references in their statements. When taken out of context, the sentences with M3S1B looked like sentences with M1S2A that aimed to provide theoretical background information; however, when the context was taken into consideration, it was seen that the communicative purpose of sentences with M3S1B was to interpret the results not to set theoretical background. Following examples (45-49) illustrate sentences with M3S1B:

(45) *Cobb (1999) showed that DDL is best suited for depth of knowledge (extending knowledge of known items) rather than breadth (adding new items), and it seems axiomatic that less advanced learners are more likely to be preoccupied with the latter* (RA7_26)

(46) ***Their line of argumentation can be extended**, and on the basis of our results, we can argue that the ability to produce original, that is, novel, ideas in general does moderately affect how students perform in a particular language learning task* (RA8_22)

(47) *That is, the explicit - inductive approach, which can lead to better accessibility of information in memory (Bruner, 1961 , 1973), helped learners with high aptitude , as measured by a memory-for-text test (Skeha, 1980 , 1982), to better retrieve grammar knowledge over time (see Hauptman, 1971; Tomlinson and Hunt, 1971)* (RA13_8)

(48) *Given the prevalence, we find it relevant to draw attention to what Currie (1998:2) has referred to as a ' tension ' that complicates the discussion of the issue of plagiarism* (RA19_13)

(49) *[this challenge relates to what Allwright (1997) refers to as the problem of sustainability in teacher research—the fact that teachers often abandon research they start doing because of the challenges it is seen to present]* (RA21_22)

M3S1 proposed in this study is roughly equivalent to the *Interpreting Results* step of *Commenting on Results* move (M4S1) in Yang and Allisons' Model. However, the statements similar to those given above (45-51) could not be identified by the definition they provided for their M4S1; because according to their model, authors referred to literature only for the purpose of comparing and contrasting their results. However, these sentences (45-51) and 57 others (a total of 64 as shown in Table 9 on page) in the corpus data were not uttered for comparative purposes, but to find support for the arguments driven from the results of the study as observed by Hopkins and Dudley-Evans (1988) and Baştürkmen (2009). Therefore, the function of referring to others' views within the context of *interpreting results* was considered a separate substep (M3S1B) in Move 3 to be able to portray all strategies used to interpret findings as well as authors' different reasons to refer to literature in Discussion sections.

Move 3, Step 2 (M3S2) – Explaining reasons for results

The purpose of the statements with M3S2 was to explain why the results were obtained as such in the present study. Similar to the *accounting for results* step of *Move 4* in Yang and Allison's Model , the authors in this study also accounted for the expected and/or unexpected results of the study mainly within the context of M3S1by discussing the possibilities that might have caused them. That is to say, after interpreting what their results suggested, they **A) provided reasons** for them as seen in the following statements (50-55).

 (50) *The finding that students who produce a high number of original solutions talk less **is probably related to** the fact that coming up with unusual solutions requires a long period of thinking time and results in a low number of solutions in general* (RA8_11)

(51) ***Another possibility is simply that*** *learners used their explicit knowledge to a greater extent in the writing test* (RA15_42)

(52) *Readers' performance in coping with inferential comprehension improves, **possibly due to** familiar(ized) subject matter allowing for more LTM contributions that help them generate more and better inferences* (RA18_14)

(53) *This discrepancy **may be explained** by the different levels of community practices in which junior and senior members engage* (RA22_25)

(54) ***It seems**, thus, **that*** *the differences found **would have to be attributed to** cross - linguistic/cross - cultural reasons* (RA24_6)

(55) *Viewing discussions as evaluations **certainly led to** greater anxiety and **probably contributed to** the students' lowered perception of their own language proficiency* (RA25_24)

In some cases, the authors **B) referred to other authors' explanations in the literature (M3S2B)** in order to account for their own results. In the following samples (56-59), for example, the authors used the explanations proposed for similar results in previous works:

(56) *Therefore, **a possible explanation** for the success of the pedagogical cycle with the less skilled listeners **is that** they were led to uncover these listening processes through guidance from the teacher and their more skilled peers (Goh , 2008)* (RA6_26)

(57) ***It is also possible that** most of the variance in task performance among students is caused by motivation (D¨ornyei, 2002), personality variables such as extraversion (Dewaele & Furnham, 2000), anxiety (MacIntyre & Gardner, 1994), and situational factors such as the interlocutor (D¨ornyei, 2002), and that among these many factors, creativity contributes to the quality of task performance only to a limited extent* (RA8_6)

(58) ***A few potential explanations have been discussed in the literature to account for** this apparent lack of need to establish a niche for one's research* (RA24_11)

(59) ***Another explanation, proposed by** Najjar (1990) (as discussed in Jogthong, 2001, p. 71), **is that** in smaller discourse communities, more typical of developing countries, authors have less pressure for publication and therefore need not be competitive for a research space* (RA24_14)

On the other hand, in some rare cases the authors **C) made evaluative comments about the explanations provided by other authors** as shown in the following examples (60-63). Since the purpose of such sentences was different than the sentences with M3S2B, they were defined as belonging to another step of Move 3, namely **M3S2C.**

(60) ***This explanation would be commensurate with*** *the increase in reported use of problem solving as listeners learned to use all information at their disposal to inference what was not understood* (RA6_38)

(61) ***Possible as it may be, this argument runs counter to*** *what research evidence has shown* (RA23_26)

(62) *In this sense, the distinction made by Samraj (2005) between established fields and emerging fields of inquiry* ***would not help explain*** *the differences between the two sets of introductions analyzed in this study* (RA24_5)

(63) *As discussed above, I do not believe this explains the BESP data (RA24_13)*

Such statements having the functions of M3S2B and M3S2C could not be classified by Yang and Allison's schematic model, because none of the moves and steps they defined specified such strategies used by the authors. This may be due to the possibility that their data coming from 20 RAs might not have involved statements with these functions or that they failed to identify reference to other authors' accounts as a different way of explaining reasons for findings, because they did not take sentence as their unit of analysis. Thus, they might have considered such statements as part of a larger segment that has the function of accounting for results in general.

HHowever, these communicative functions were observed by Baştürkmen (2009) in 10 RA discussions that came from a single journal, i.e. Language Teaching Research Journal. She also defined *providing alternative explanations for the same results* as one way of accounting for results signaled by phrases like *Another reason/A related explanation may be that…. S*imilarly, the data of the present study involved instances of offering additional explanations both in M3S2A and M3S2B. Such statements involved phrases like *Another possibility is simply that,* (51), *It is also possible that..* (57), *A few potential explanations have been discussed in the literature to account for …* (58), *Another explanation, proposed by…. , is that….(59).* Therefore, these observations are congruent with the observations made by Baştürkmen.

Move 3, Step 3 (M3S3) – Comparing to the results of previous work

This move was used to compare the results of the present study to those of similar studies. In other words, by referring to literature, the authors stated **A) whether or not the results support the findings/claims/hypotheses of other authors (M3S3A)** as seen in the following sentences (64-70):

(64) *Observations made during the experiment **provide some evidence** (albeit anecdotal) pointing towards anticipatory processes **similar to those proposed by** Hoey (2005) and McDonald and Shillcock (2003a, 2003b)* (RA1_21)

(65) *The higher rate of occurrence for reformulation **replicates the findings of some of the previous research** in both classroom and dyadic contexts that have shown that recasts are more dominant than other types of interactional feedback (Ellis et al.., 2001; Lyster & Ranta, 1997; Nabei & Swain, 2002; Oliver, 1995, 2000; Panova & Lyster, 2002)* (RA3_8)

(66) *The results of this third research question **confirm recent studies suggesting that** students can indeed improve their L2 oral proficiency during a one - semester study - abroad program (Magnan & Back, 2007; Segalowitz & Freed, 2004)* (RA10_24)

(67) *The ability of morphology to predict reading comprehension over and above phonology is **in line with previou**s L1 reading **research** (e. g., Carlisle, 1995, Deacon & Kirby, 2004; Nagy et al., 2003, 2006; Singson et al., 2000), most of which investigated young L1 (mostly English monolingual) children* (RA11_7)

(68) ***This result is also consistent with previous L1 studies** that reported morphology as a significant predictor of reading comprehension over and beyond vocabulary knowledge (e. g. ,Carlisle, 2000; Katz , 2004; Singson et al., 2000)* (RA11_26)

(69) *In particular, **research was contrasted with** reflective practice; **a similar distinction is made by** Cochran - Smith and Lytle (1999), who argue that teacher r*esearch goes beyond the kind of thoughtful teaching that reflective practice involves (RA21_16)

(70) ***The present study confirms findings from previous research concerning** why teachers error correct the way they do* (RA34_1)

This step is equivalent to Yang and Allison's M4S2. As a matter of fact, comparing findings to those of previous studies, is the most frequently reported reason for authors to refer to literature in their discussions. Almost all of the Discussion models reviewed for the present study have a reference to literature step for this specific purpose (e.g. Dubley-Evans, 1994; Nwogu, 1997; Posteguillo, 1999; Peacock, 2002; Yang and Allison, 2003; Kanoksilapatham, 2005). However, these models do not mention that the authors may also B) **provide information about the results of these studies (M3S3B)** and **C) make**

comments on their results (M3S3C) after stating whether or not the current results are in congruent with theirs. In other words, using these steps the authors report and interpret what the results of other studies with similar findings suggest; provide explanations for these findings; and criticize them by referring to a gap in their methodology in an attempt to explain differences in findings. Such uses of M3S3B (71-72) and M3S3C (73-74) are illustrated below:

(71) ***Both studies investigated*** *older (Grades 8 and 10) L1 Dutch speakers* ***and reported that*** *metacognition played an important role in L2 English reading comprehension* (RA11_31)

(72) ***However, studies by*** *Allwright (1975) , Chaudron (1988) , and Lyster (1998)* ***have shown that*** *this implicit strategy is less effective than more explicit strategies as the students may have difficulties recognizing the strategy as a corrective feedback* (RA20_24)

(73) *The high amount of empirical support for the concreteness superiority effect* ***from previous research is probably related*** *in part to the high number of studies that tested recall a short period of time after the experiment* (RA2_8)

(74**)** ***In addition, neither*** *Schoonen et al.* ***nor*** *van Gelderen et al.* ***reported*** *the reliabilities or validity evidence of their metacognition measures, which makes it difficult to determine the quality of the measurement instruments used in these studies* (RA11_37)

<u>Move 3, Step 4 (M3S4) – Evaluating the results</u>

Using this step, the authors of the RAs within the corpus of the study evaluated the results of their own studies by making a subjective comment on them as seen in the following sentences (75-77):

(75) *These mixed results are **somewhat surprising*** (RA4_23)

(76) *As we said above, **this result should be interpreted with some caution**, since the word estimate, was coded as a hedge even though it appeared in the assignment brief (RA28_17)*

(77)   *Teachers' suggestions that examinations were a barrier to task-based approaches **were not fully convincing to me** in that high-stakes public examinations in Hong Kong have been moving in a more task-based direction for some years* (RA31_17)

<u>Obligatory, Conventional, and Optional M3 categories and their frequency of occurrence in the quantitative and qualitative RAs:</u>

The number and percentages of the RAs as well as the number of sentences in which they were observed are displayed in Table 8. Their frequency of occurrence at step level, on the other hand, is shown in Table 9.

Table 8. Frequency of Occurrence of M3 at Substep Level

	N (%) of RAs containing M3 Steps in the Corpus (N=36)	QUAN RAs (N=18)			QUAL RAs (N=18)			X^2
		N (%) of RAs containing M3 steps	f in QUAN RAs	% of all QUAN Moves	N (%) of RAs containing M3steps	f in QUAL RAs	% of all QUAL Moves	
M3S1A	35 (97.22) **	18 (100)*	119	9.74	17 (94.44)**	172	16.21	1.00 (df=35)
M3S1B	22 (61.11)**	12 (66.67) **	22	1.80	10 (55.56)	42	3.96	.733 (df=22)
M3S2A	24 (66.67)**	14 (77.78)**	64	5.24	10 (55.56)	30	2.92	.711 (df=26)
M3S2B	14 (38.89)	9 (50)	38	3.11	5 (27.78)	15	1.41	.305 (df=14)
M3S2C	5 (13.89)	3 (16.67)	3	0.24	2 (11.11)	5	0.47	1.00 (df=14)
M3S3A	30 (83.33)**	18 (100)*	75	6.14	12 (66.67)**	35	3.20	.023 (df=28)
M3S3B	18 (50)	11 (61.11)**	37	3.03	7 (38.89)	14	1.32	.641 (df=17)
M3S3C	9 (25)	7 (38.89)	19	1.55	2 (11.11)	2	0.19	.121 (df=9)
M3S4	25 (69.44)**	15 (83.33)**	25	2.05	10 (55.56)	17	1.60	.304 (df=24)

*obligatory, ** conventional

Table 9. Frequency of Occurrence of M3 at Step Level

	N (%) of RAs containing M3 Steps in the Corpus (N=36)	QUAN RAs (N=18)			QUAL RAs (N=18)		
		N (%) of RAs containing M3 steps	f in QUAN RAs	% of all QUAN Moves	N (%) of RAs containing M3 steps	f in QUAL RAs	% of all QUAL Moves
M3S1	35 (97.22)**	18 (100)*	141	11.54	17 (94.44)**	214	20.17
M3S2	25 (69.44)**	14 (77.78) **	105	8.59	11 (61.11)**	50	4.81
M3S3	31 (86.11)**	18 (100)*	131	12.76	13 (72.22)**	51	6.31
M3S4	25 (69.44) **	15 (83.33)**	25	2.04	10 (55.56)	17	1.60

*obligatory, ** conventional

As the values in Table 9 revealed, two *obligatory* M3 steps were identified for the quantitative RAs: **M3S1** that included the authors' **interpretations of their results**; and **M3S3** through which the results were **compared to the previously reported results of other studies**. These two steps were realized in all quantitative RAs through their first substeps, namely **M3S1A** and **M3S3A** (Table 8). In the quantitative subcorpus of the study, the authors **referred to literature to interpret their results with established knowledge** using **M3S1B** in 66.67% (N=12) of the RAs, and to compare their findings using **M3S3B** in 61.11% (N=11) of the RAs. These substeps were, hence, *conventional* to use in quantitative research reports. To **make evaluative comments** on the previous studies the results of which were reported for comparative purposes (**M3S3C**), however, was an *optional* purpose to achieve for the authors.

On the other hand, **M3S1A (stating what the results suggest and making claims/ arguments based on the findings of the study)** and **M3S3A (stating whether or not the results support the findings/claims/hypotheses of other authors)** that appeared in 94.44% (N=17) and 66.67% (N=12) of the qualitative RAs respectively were considered *conventional* to use in this subcorpus (Table 8). In fact, all articles with both quantitative and qualitative

approaches to research were expected to involve the interpretation of results either through M3S1A and/or M3S1B in their Discussion sections in line with the acknowledged function of this section. However, in the qualitative subcorpus of the study, the authors of one of the RAs (RA35) did not interpret what was suggested by the individual findings in the Discussion section; instead, they interpreted and accounted for their findings in the Results (Findings) section of their RA. On the other hand, in the Discussion section of this particular article, they frequently used the steps of deduction move (M6) which were generally observed to suggest classroom and research implications after the discussion of results. Table 8 also shows that realizing M3S1 and M3S3 by referring to established knowledge or other authors' views through **M3S1B (interpreting results by referring to other authors' views) and M3S3B (giving information about the results of other studies**) was *optional* to use in qualitative research as they were observed in 55.56% (N=10) and 38.89% (N=7) of the RAs respectively. **M3S3C (making comments about other studies)**, however, was identified only in 2 (11.11%) of the qualitative RAs.

As for the other M3 steps, **M3S4-evaluating the results** observed in 83.33% (N=15) of the quantitative RAs and **M3S2- explaining the results** by providing reasons through **M3S2A** observed in 77.78% (N=14) of the quantitative RAs were identified as *conventional* for this research type. In half of the RAs (N= 9) within the quantitative subcorpus, the authors referred to other authors' explanations to account for their results **(M3S2B)**. These steps and substeps that were observed less frequently in the qualitative RAs were found to be *optional* to use unlike the quantitative RAs.

In summary, it was *obligatory* to use M3S1-Interpreting Results and M3S3-Comparing Results; but *conventional* to use M3S2-Explaining Reasons for results and M3S4-Evaluating Results in the quantitative RAs. In the qualitative RAs, on the other hand, it was

conventional to interpret and compare; but *optional* to explain reasons and evaluate the results.

As for the number of sentences that included the M3 substeps, the frequency count results revealed 291 occurrences of interpretation through M3S1A in the corpus of the study. 119 of these appeared in the quantitative RAs (Table 8) comprising 9.74% of all quantitative moves (M=9.44, SD=5.54) while 172 appeared in the qualitative RAs comprising 16.21% of all qualitative moves (M=17.01, SD=13.99) identified in the study (Figure 4).

Figure 4. Percentage of quan and qual moves that are made up of M3 Steps & Ss
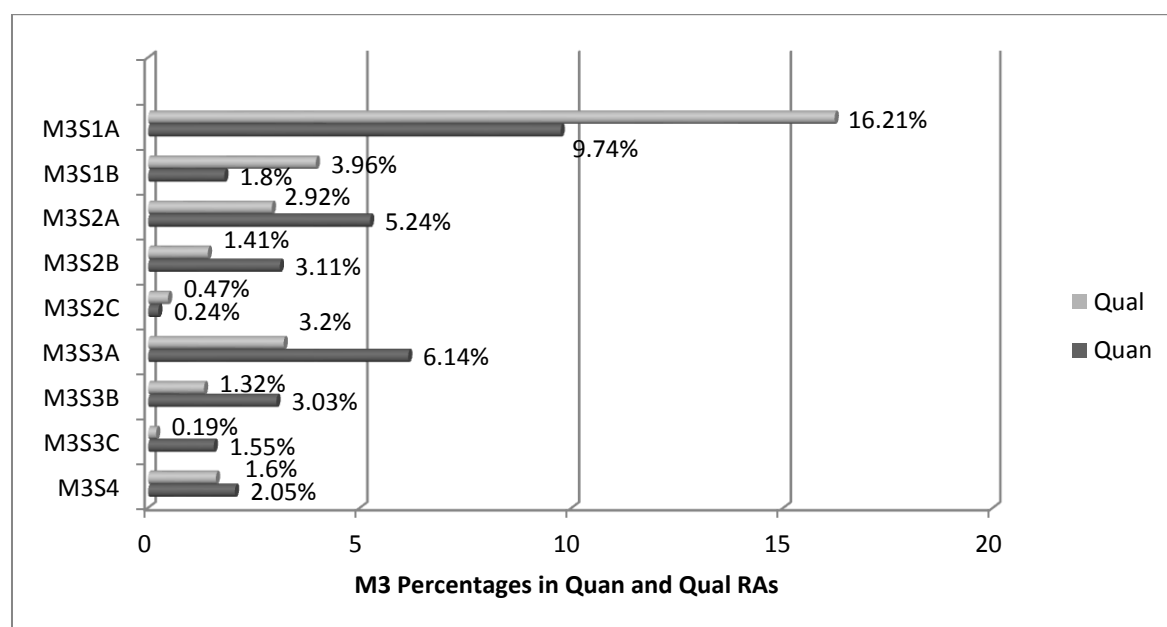
The other obligatory move for the quantitative RAs, M3S3A, however, appeared in a total of 110 sentences, 75 in the quantitative and 35 in the qualitative subcorpus, comprising 6.14% of all quantitative (M=6.12, SD=3.67) and 3.2% of all qualitative (M=3.22, SD=3.05) moves. This difference was found to be significant with a X^2 value of *(28, N=36)=34.00*, *p=.023* (p< .05) confirming that authors compare their findings to those of other studies more frequently in quantitative research. The study revealed no significant difference in the occurrence of other M3 categories in both types of research articles.

Figure 5, on the other hand, shows the percentages of M3 categories at step level. The percentages in this figure better shows that the interpretation of results (M3S1) is the most frequently used M3 step in both quantitative (20.17%) and qualitative (11.54%) research. It is followed by comparing results (M3S3) in quantitative (10.72%) and accounting for results (M3S2) in qualitative research (4.81%).

Figure 5. Percentage of quan and qual moves that are made up of M3 Steps
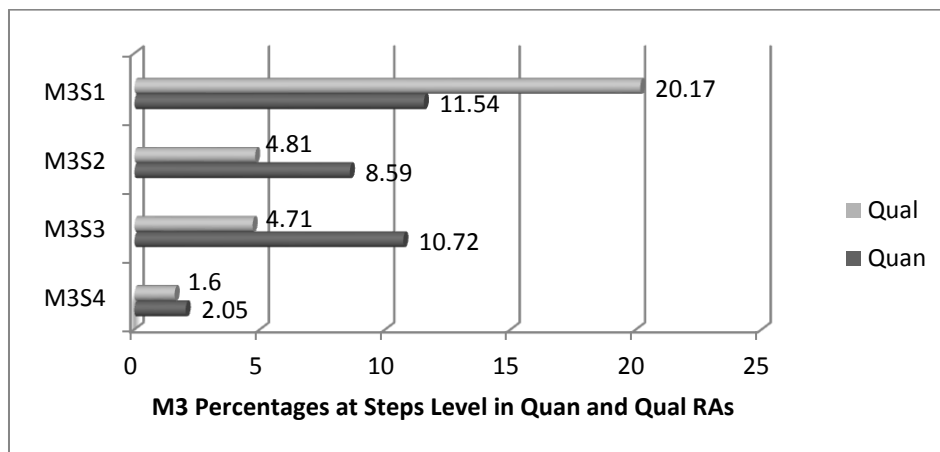
Move 4 (M4) – Summarizing the Study

The authors used Move 4 generally in conclusion sections to make a summary of the study. It corresponds to Yang and Allison's Move 5 that they defined as summarizing the main points from the perspective of the study. However, this definition proved problematic during the initial coding of the data based on this model, because it did not clarify what those main points involved. Therefore, the definition was revised for an easier identification of the statements that summarize **what has been done in the study (M4S1)** and **what has been suggested by the study (M4S2).** In other words, **Move 4, Step 1- M4S1** provided a brief methodological summary reflecting the perspective of the study while **Move 4, Step 2- M4S2** summarized the conclusions that the authors reached having interpreted the main results of the

study. The uses of M4 steps have been illustrated in the sentences (78-81 for Step 1; 82-85 for Step 2) provided below:

(78) ***This study has combined these approaches to investigate a topic*** *that in recent years has received little attention the impact of learner errors* (RA1_C67)

(79) ***Our overarching aims were to investigate*** *whether intervention via a listening strategy program was worthwhile with a population of students who appear to find listening at the lower-intermediate level one of the most difficult skills and who, past research suggests, appear to lack both the tools and affective disposition to overcome their problems* (RA4_C61)

(80) ***In this study, we have provided an analysis of the emic perspectives*** *of NNES students and their advisors in three different disciplines* (RA22_C46)

(81) ***The present paper reports an analysis of*** *the discourse features of six teaching practicum reflective reports written by pre-service student teachers of English in Hong Kong* (RA32_26)

(82) ***In summary***, *the results of this study confirm the role of salience and explicitness as important characteristics of effective feedback in dyadic student - teacher interaction* (RA3_C66)

(83) ***We can conclude from the findings that*** *the intervention was , essentially , beneficial both in terms of improving listening proficiency and raising the students' self–efficacy* (RA4_C62)

(84) ***As a result of*** *these allowable linguistic differences and similarities in the written data,* ***it may be concluded overall that*** *English and Italian writers are in a position to*

provide their opinions with a social expendibility, as defined by the socio - rhetorical action of the genre for specific communities (RA27_C21)

(85) ***In conclusion, this paper bolsters claims that*** *we need context-sensitive teaching methods or what has been termed here as situated task-based approaches, in which culture, setting and teachers' existing beliefs, values and practices interact with the principles of task based teaching* (RA31_C35)

When read out of context, sentences above (78-81) looked quite similar to the sentences with M1S1 that provided a review of methodological information regarding the study. However, M1S1 was used generally in the beginning of Discussion sections for the purpose of setting the context in which the results would be interpreted and evaluated. The purpose of the sentences with M4, however, was to summarize the discussion section in relation to what was done (M4S1) and what was found (M4S2) before concluding the study.

Obligatory, Conventional, and Optional M4 categories and their frequency of occurrence in the quantitative and qualitative RAs:

The frequency of occurrence of M4 steps in the corpus of the study is shown in Table 10:

Table 10. Frequency of Occurrence of Move 4 Steps

	N (%) of RAs containing M4 Steps in the Corpus (N=36)	QUAN RAs (N=18)			QUAL RAs (N=18)			
		N (%) of RAs containing M4steps	f in QUAN RAs	% of all QUAN Moves	N (%) of RAs containing M4 steps	f in QUAL RAs	% of all QUAL Moves	X^2
M4S1	17 (47.22)	9 (50)	20	1.64	8 (44.44)	11	1.04	.740 (df=18)
M4S2	29 (80.56)**	14 (77.78)**	37	3.03	15 (83.33) **	55	5.18	.530 (df=28)

*obligatory, ** conventional

Figure 6. Percentage of quan and qual moves that are made up of M4 Steps
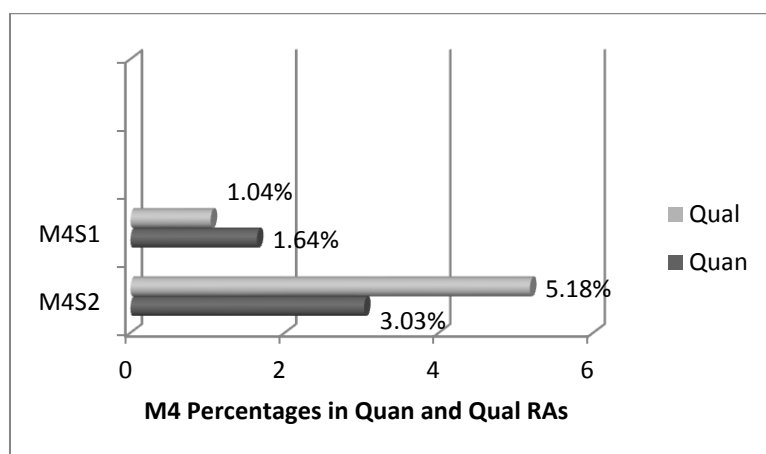
According to the values in Table 10 and Figure 6, the authors *conventionally* used M4S2, 55 times in 83 % (N=15) of the qualitative RAs, and 37 times in 78 % (N=14) of the quantitative RAs. This step comprised 5.18% of all qualitative moves and 3.03% of quantitative moves. On the other hand, M4S1 identified in fewer sentences (20 times in quantitative RAs & 11times in qualitative RAs) was an *optional* step to realize M4 in both subcorpora. These findings confirm that quantitative RAs provided more methodological summary while qualitative RAs provided more summary of the findings.

Move 5 (M5) – Evaluating the study

Another purpose that the authors aimed to achieve in the RA Discussion sections was found to be the *evaluation of the study*. The present study identified three steps through which the authors evaluated their studies:

Move 5, Step 1 (M5S1) – Acknowledging limitations

The authors used this step in order to acknowledge the limitations of their study caused by or inherent in its methodology. They speculated on unexpected factors that might have affected the results obtained; and indicated what was left out of the scope of the study. Thus, they justified the need for future work by pointing to the issues that could not be addressed by the study.

These functions of M5S1 are illustrated in the following sentences (86-91):

(86-87) *It should be emphasized, however, that the experimental power for distinguishing between PI and TI was potentially too low to detect any differential effects,* **which limits the usefulness of the current study in** *comparing the effects of different types of instruction on language development in heritage speakers. The lack of statistical power leads us to some additional methodological limitations of the current study that should be considered* (RA5_43-44)

(88) *Finally,* **the study has not been able to address** *the cognitive processes that underlie linguistic development* (RA5_51)

(89) *The use of a single topic for all three writing tasks would not have removed this year–task compound because the writers would have been affected by their previous solutions to the task, and the recurrence of the task may have caused a decrease in motivation* (RA16_31)

(90) *Due to the nature of the study and time constraints, **the study is limited in its scope** as it investigates the beliefs and feedback practices of a small sample size of only three teachers of English language, in three Singapore neighborhood schools* (RA20_27)

(91) ***The results cannot be generalized to** other writing contexts due to the small number of participants involved and their almost identical cultural backgrounds* (RA23_C60)

Move 5, Step 2- Indicating the significance of the study

Having used this step, the authors evaluated their studies considering their significance and contribution to the field of research as shown in the following samples (92-97):

(92) ***Our study contributes to** learner strategy theory in general by exploring the link among strategy use, self - efficacy, and attributions* (RA4_44)

(93) ***This study makes a number of** important **methodological contributions*** (RA6_43)

(94) ***The current study is different from previous** written CF **studies in that** only one linguistic feature was targeted for the provision of CF and the tests developed measured students ' written accuracy alone* (RA15_4)

(95) *The conceptions of research highlighted here **contribute to** an understanding of why research for many teachers can seem to be an irrelevant and unfeasible activity* (RA21_20)

(96)   ***The research reported here offers insights into*** *the change processes experienced by beginning teachers (RA29_C79)*

(97)   ***This study provides a useful reminder about*** *the importance of systematic evaluation for syllabus designers and curriculum developers who are creating task-based courses (see, e.g., Watson Todd, 2006) (RA35_20)*

Move 5, Step 3 (M5S3) – Justifying the methodology

This step was mainly used for the purpose of justifying the methodology of the study. Unlike the first step of M5 through which the authors acknowledged their methodological weaknesses, using M5S3 they provided reasons for their methodological decisions and asserted evaluative comments to recognize their strengths. The following sentences (98-101) reveal these functions identified in the corpus data:

(98-99) *We ensured that the testing procedure did not bias the intervention students by implementing a listening - test type that was not practiced during the strategy instruction program. By doing so, we believe that we also provide evidence that students were able to transfer their strategic behavior from the tasks they engaged in during the instruction to a different task in the listening tests (RA4_41-42)*

 (100)   *That said , all attempts have been made to minimize the effects of the limitations of the study to increase the validity , reliability , authenticity , as well as the ethics of the study (RA20_34)*

(101)   *I employed both a textual analysis method and retrospective interviews to present a more accurate picture of the writers ' attitudes toward reviewers ' stances and comments (RA23_36)*

109

This move is almost equivalent to Move 6 identified by Yang and Allison who also presented three different possibilities, namely *indicating the limitations, indicating the significance* and *evaluating the methodology,* to evaluate the study. However, attempts to distinguish *indicating limitations* and *evaluating the methodology* were problematic in cases where the authors evaluated the study in terms of its limitations. In other words, evaluation included the assessment of the study both by its weaknesses and strengths mainly caused by its methodology. Therefore instead of the term *evaluate*, the term *justify* was suggested and hence, statements that evaluated the study by its methodological weaknesses were classified by the step of *limitations* while those that evaluated the study by its methodological strength were classified by the step of *justifying the methodology.*

Obligatory, Conventional, and Optional M5categories and their frequency of occurrence in the quantitative and qualitative RAs:

The frequency of occurrence of M5 steps are displayed in Table 11.

Table 11. Frequency of Occurrence of Move 5 Steps

	N (%) of RAs containing M5 Steps in the Corpus (N=36)	QUAN RAs (N=18)			QUAL RAs (N=18)			
		N (%) of RAs containing M5steps	f in QUAN RAs	% of all QUAN Moves	N (%) of RAs containing M5 steps	f in QUAL RAs	% of all QUAL Moves	X^2
M5S1	23 (63.89)**	15 (83.33)**	58	4.75	8 (44.44)	25	2.26	.035 (df=23)
M5S2	16 (44.44)	9 (50)	19	1.55	7 (38.89)	17	1.60	.738 (df=16)
M5S3	9 (25%)	7 (38.89%)	21	1.72	2 (11.11%)	4	0.38	.121 (df=9)

*obligatory, ** conventional

Figure 7. Percentage of quan and qual moves that are made up of M5 Steps
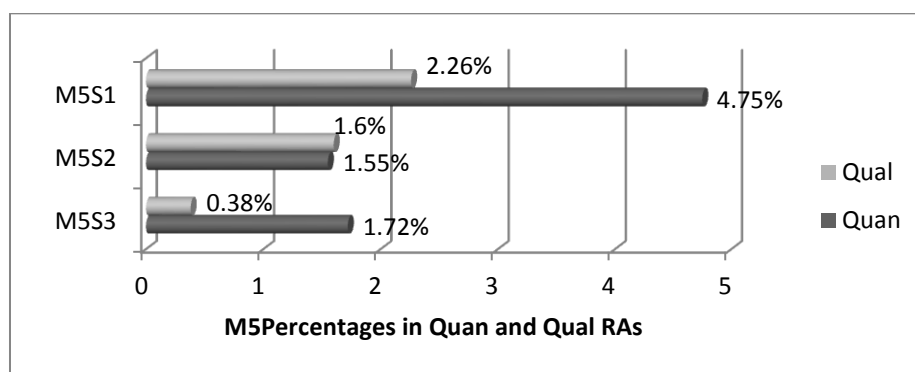
As the Table 11 shows, among all steps of M5, **M5S1- acknowledging limitations** identified in 58 sentences in 83% (N=15) of the quantitative RAs was the only M5 step used *conventionally* by the authors to evaluate their study in the quantitative corpus. This step made up 4.75% of all moves identified in the quantitative subcorpus (Figure 7). On the other hand, it appeared 25 times in 44% (N=8) of the qualitative RAs; therefore, it was considered *optional* in the qualitative RAs in which it accounted for 2.26% of all moves. This differences was found to be significant with a value of *(23, N=36)=9.926, p=.035*, (p<.05) confirming that in quantitative research the authors mention their limitations significantly more often than they do in qualitative research. This finding seems to be congruent with a previous finding of the present study regarding authors' tendency to provide significantly more methodological information in quantitative research.

As for the other M5 steps, M5S1 was followed by M5S2 that was identified 19 times in half (N=9) of the quantitative RAs and 17 times in 39% (N=7) of the qualitative RAs. This finding shows that evaluating the study by highlighting its significance or contribution to the area of research is an optional move to use in RA discussions. Similarly, the final step of M5S3- justifying the methodology- was a step observed only in a quarter (N=9) of the RAs (7 quantitative and 2 qualitative) that formed the corpus of the study. The figures reveal that the

authors of the qualitative RAs rarely needed to justify their methodology in the Discussion section of their RAs.

<u>Move 6 (M6) – Drawing Implications</u>

The move analysis of the study revealed two types of implications drawn based on the interpretation of the results by the authors in RA discussions: **Implications for future research, M6S1,** and **pedagogic implications, M6S2.**

<u>Move 6, Step 1 (M6S1) – Implications for future research</u>

Using this step, the authors of the RAs recommended further research and made recommendations for prospective researchers. They achieved this purpose by **A) stating and justifying the need for future research (M6S1A).** In other words, the authors indicated the need for future studies in the same vein by providing reasons and emphasizing the contribution they would make or the gap they would fill in the related literature. The authors also raised questions to be investigated and pointed to the issues that needed to be considered in future research; and in some cases, they referred to some previous studies to illustrate the suggested issues. The sentences that were observed to have these functions are provided below (102-110):

(102) *Perhaps more importantly, **these findings also highlight the need for future investigations into** the role of input - based factors, especially in the learning of languages other than English* (RA9_C70)

(103) *Validation of this idea **would be a suitable topic for a** follow - up instructional treatment **study** (*RA11_64)

(104-105) *Such research will allow researchers to* *empirically identify the latent construct(s) of interest and possibly effective test items. Muse's study (2005) can be thought of as such an exemplary investigation of L1 morphology* (RA11_C70-71)

 (106) *Further research is needed to examine* *the effectiveness of the selective approach to CF recommended earlier* (RA15_56)

 (107) *Issues requiring further investigation include* *the extent to which teachers may use examination requirements as a pretext or justification for the kind of approach that they personally favour; is most accepted in their social setting; or is most practical to implement* (RA31_23)

 (108-110) *Future research in the following areas can help* *L2 teachers design and implement task-based courses: How effective are task-based courses in promoting L2 learning? How can task-based language teaching be adapted to optimize learning outcomes in a wide variety of instructional contexts* (RA35_C37-39)

The authors also achieved M6S1 by **B) making methodological suggestions for future research (M6S1B)** having considered the methodological problems they experienced and the limitations of their studies; and occasionally referred to other authors' views to seek support for the suggestions they made. These functions are illustrated in the following sentences (111-116):

(111) *To explore the hypotheses outlined above, **future research might use** an eye tracking **methodology to** generate information regarding the point at which the regressions occur, the location to which the reader regresses, and the duration of any subsequent fixations* (RA1_54)

(112-114) *The pedagogical cycle under investigation in the current study **could be** **further enriched by** adding a "bottom - up" component to the third phase of the cycle. Not only could this stimulate interest in the third listen, but this step would also lead to greater noticing of words , expressions , and syntactic structures. Such a "noticing activity" is also advocated by Richards (2005) for purposes of language awareness leading to acquisition activities* (RA6_C55, 58-59).

(115) ***In future research**, these variables **should be** kept the same as much as possible* (RA34_C30)

(116) ***In future research**, participating teachers **can be** recruited from the same type of educational institutions so that variables between teachers can be kept as constant as possible* (RA34_C32)

These steps identified as M6S1A and M6S1B are similar to Yang and Allison's M7S1 and M7S2

Move 6, Step 2 (M6S2) – Pedagogic implications

This move was realized in two ways by the authors of the corpus data: Having interpreted what is suggested by the findings of the study, the authors **deducted practical classroom implications -M6S2A-** and made guiding suggestions for teaching. Thus, before concluding the article, they paraphrased their findings in terms of what could be done by language teachers, language learners, curriculum and test developers and language programs to improve teaching as observed in the following sentences (117-121).

(117) *Nevertheless , **our results suggest that language teachers may** resort to writing tasks that incorporate newly taught words in order to enhance students' retention* (RA2_15)

(118) *Study-abroad program directors and instructors can then continue to promote integrative motivation through similar activities within the study abroad environment* (RA10_61)

(119) *At the same time, **instructors can provide students with** frequent and sustained opportunities to use the L2 outside of class through participation in a language exchange program* (RA10_64)

(120) *In particular, the role of grammar instruction in task-based language teaching should be highlighted so that teachers understand the difference between focus on form and focus on forms and learners can discuss their attitudes and expectations about explicit grammar instruction* (RA35_26)

(121) *In testing, then, reading topics should be selected with a view to minimizing interference from the text, so as to have a fair evaluation of test takers' inferential reading performance* (RA18_C36)

While suggesting the classroom implications, the authors sometimes **B) referred to literature to seek support for the implication to be drawn** -**M6S2B.** Sentences identified with this function are illustrated below (122-127):

(122) *To enhance students' perceptions of competence, teachers should use individual criteria to evaluate and reward students' performance (Ames, 1992) and to provide*

students with verbal feedback informing them that their performance is competent (Valler and and Reid, 1984) (RA14_29)

(123-124) *Sullivan and Lindgren (2006) argue that replay provides students "with appropriate input" through the use of the student's own output (p . 205). In the TESOL setting, such input could include aspects of English grammar, vocabulary, discourse, and style, which may be taken up in class and discussed* (RA16_42-43)

(125) *In particular, as Carson (1998) and Leki (2001) pointed out, in linguistically and culturally diverse learning situations, it seems important for instructors to appreciate the diverse language and cultural backgrounds of the students, closely follow their learning processes , and to intervene in them if and when needed.* (RA25_C54)

(126) *A potentially effective way to socialize novice genre users into such discourse norms could be through engaging them in conducting ethnographic study of expert genre users' practices through observations and interviews (see Johns, 1997; Lave and Wenger, 1991), and/or through joint lesson study analysis and subsequent reflective discussions as a key teacher education activity (see Lewis et al., 2004; Marton and Lo, 200)* (RA32_25)

(127) *Lewis (2000b) listed several classroom activities for teaching collocations* (RA17_53)

Move 6 defined in this study can be considered a revised form of *Move 7-Deductions from the research* suggested by Yang and Allison. According to their model, authors *make suggestions (M7S1)* to solve the problems they encountered in their research; *make recommendations for further studies (M7S2)*; and *draw pedagogic implications (M7S3).*

However, since two of these steps, M7S1 and M7S2, concerned research, their definition overlapped and proved problematic during the move analysis. That is to say, almost all of the sentences that included M7S1 were observed to have M7S2 because solutions to the problems were uttered as recommendations to be considered by prospective researchers in future studies. Therefore, these steps were redefined to clearly indicate authors' purpose to state the need for making further studies in some suggested areas considering the methodological recommendations drawn from the study.

The revised steps of the new coding scheme also clearly pointed to the two main types of implications, namely for research and classroom, drawn by the authors. The research implications were intended for expert (e.g. university academics) or novice (e.g. graduate students) researchers who may be interested in making similar studies within the same field of research; whereas classroom implications were intended for practicing teachers, testers, and program and curriculum developers. Since the subjects and the purpose of these recommendations are different, they can be considered as different steps of *drawing implications move*.

In addition, the new steps also revealed authors' tendency to refer to literature to seek support for the implications drawn, which was not mentioned at all in Yang and Allison's Model.

Obligatory, Conventional, and Optional M6categories and their frequency of occurrence in the quantitative and qualitative RAs:

The frequency counts of the RAs and the sentences that included M6 steps and substeps are provided in the following tables (Table 12 & 13):

Table 12. Frequency of Occurrence of Move 6 at Substep Level

	N (%) of RAs containing M6 Steps in the Corpus (N=36)	QUAN RAs (N=18)			QUAL RAs (N=18)			X^2
		N (%) of RAs containing M6steps	f in QUAN RAs	% of all QUAN Moves	N (%) of RAs containing M6 steps	f in QUAL RAs	% of all QUAL Moves	
M6S1A	30 (83.33)**	17 (94.44)**	68	5.56	15 (83.33)**	51	4.81	.795 (df= 31)
M6S1B	22(61.11)**	18 (100)*	74	6.05	9 (50)	28	2.64	.007 (df= 25)
M6S2A	33 (91.6)**	16 (88.89)**	118	9.65	17 (94.44)**	165	15.55	1.00 (df= 33)
M6S2B	27 (75)**	11 (61.11) **	47	3.85	16 (88.89)**	45	4.24	.194 (df= 23)

*obligatory, ** conventional

Table 13. Frequency of Occurrence of Move 6 at Step Level

	N (%) of RAs containing M6 Steps in the Corpus (N=36)	QUAN RAs (N=18)			QUAL RAs (N=18)		
		N (%) of RAs containing M6steps	f in QUAN RAs	% of all QUAN Moves	N (%) of RAs containing M6 steps	f in QUAL RAs	% of all QUAL Moves
M6S1	33 (91.67) **	18 (100)*	142	11.62	17 (94.44)**	79	7.44
M6S2	33 (91.67) **	17 (94.44)**	165	13.50	17 (94.44)**	210	19.79

*obligatory, ** conventional

Figure 8. Percentage of quan and qual moves that are made up of M6 Steps
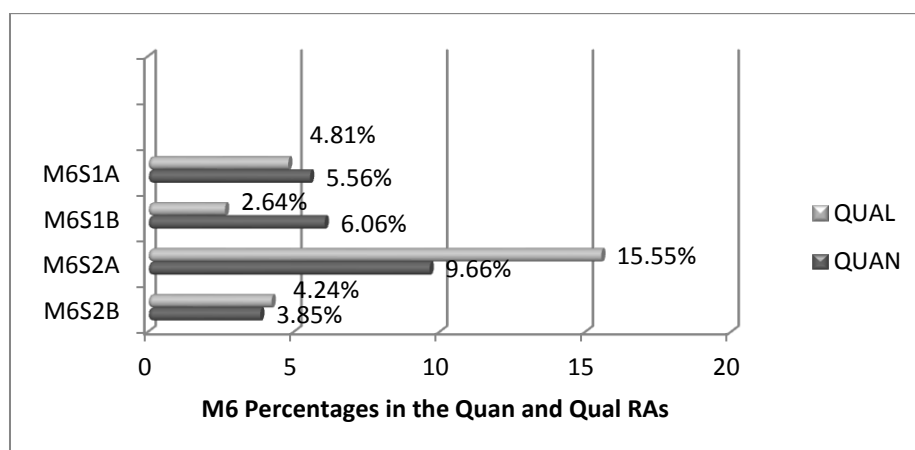
As shown in Table 12, **M6S1B** through which the authors offered methodological suggestions for the future studies was an *obligatory* move identified 74 times in all of the quantitative RAs, which made up the 6% of the quantitative moves. However, this step was observed in half as many of the qualitative RAs as an *optional* step and included only 29 occurrences which made up 2.6% of all qualitative moves. This difference was found to be significant with Fisher's exact p value of .007 (p< .05), *(25, N=36)=32.00*, suggesting that giving methodological recommendations is a move related to the research type. In quantitative RAs it is used significantly more often by the authors.

While the authors of the qualitative RAs *optionally* made methodological suggestions, they *conventionally* stated and justified the need for further studies **(M6S1A)** in the concluding sections like the authors of the quantitative RAs. This function was identified 68 times in the 94.44% (N=17) of the quantitative RAs and 51 times in the 83.33% (N=15) of the qualitative RAs.

The second step of M6 through which the authors deducted classroom implications with or without reference to other authors (**M6S2B** and **M6S2A**), was identified as *conventional* to use in both quantitative and qualitative RAs (Table 12). Although M6 steps

(M6S1 and M6S2) were observed in almost the same number of RAs in both subcorpora, sentences with M6S2 (165 in quan RAs and 210 in qual RAs) outnumbered those with M6S1 (142 in quan RAs and 79 in qual RAs) in both types of RAs (Table 13). These numbers show that the authors of both quantitative and qualitative RAs provided more classroom implications than research implications in their discussions. The findings also revealed that drawing classroom implications comprised 13.5% of all quantitative moves versus 19.79% of qualitative moves; whereas drawing research implications comprised 11.6% of all quantitative moves versus 7.44% of qualitative moves. What these findings suggest is that the qualitative RAs provided more classroom implications than the quantitative RAs while the quantitative RAs provided more research implications than the qualitative RAs.

Percentages of occurrence at Move Level

Having discussed the distribution of all move-step categories by research types, the percentages comprised by 6 Moves including their subcategories in both subcorpora was calculated to see the most common purposes of authors in discussion sections. These percentages are shown in the following figure:

Figure 9. Distribution of 6 moves in quantitative and qualitative RAs
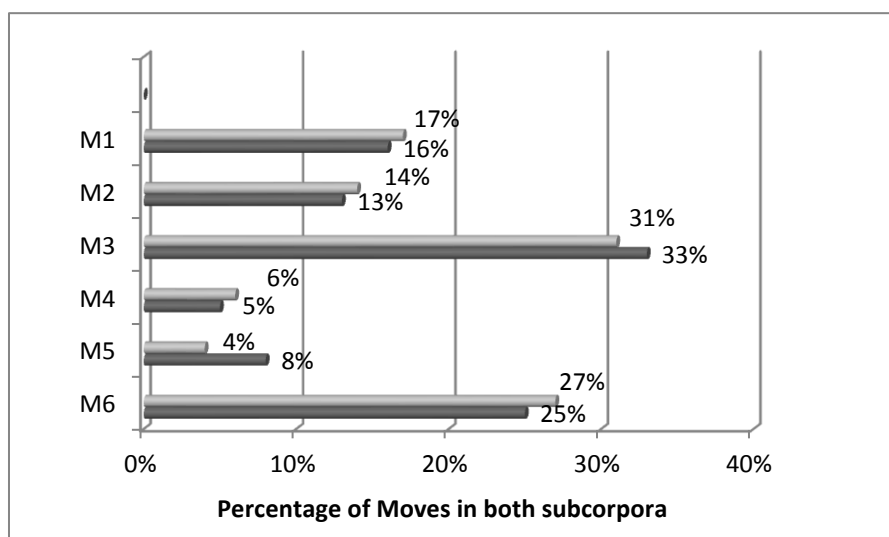
As seen in Figure 9, ***discussing results (M3)*** using either one(s) of interpreting, explaining reasons, comparing, and evaluating is the main purpose of authors in Discussion sections. M3 with all its subcategories made up 33% of all quantitative and 31% of all qualitative moves identified in the study. It was followed by the deduction move used to ***draw implications (M6)*** primarily for classroom and future research. M6 made up 27% of all qualitative moves and 25% of quantitative moves, and was followed by ***background information move (M1)*** that comprised 17% and 16% of qualitative and quantitative moves respectively.

Reporting results (M2), however, was the fourth common move used in both subcorpora. It made up 14% of all qualitative and 13% of quantitative moves. In quantitative corpus, it was followed by the moves of ***evaluating (M5)*** and ***summarizing the study (M4)***. It should be noted that the difference between the distribution percentages of the moves in both subcorpora was the highest in **M5.** It made up 8% of all quantitative moves; but 4% of all qualitative moves. In fact, in qualitative research, ***evaluating the study*** was the least common purpose shared by the authors.

4.1.2 Move Cycles

In order to answer the Research Question 1.4, move sequences observed with a minimum frequency of 12 times in at least 6, one third of 18, RAs were tabulated and shown in Tables provided in Appendices I-J & K. These numbers were decided arbitrarily considering the size of the corpus. As the size of two subcorpora is not equal in terms of the number of units analyzed and of moves identified, the observed frequency of move cycles was normalized for comparison using the equation $\frac{\text{observed frequency} * 1000}{\text{number of moves identifed}}$. For example, the sequence of M2-M3 that occurred 72 times in the quantitative and 52 times in the qualitative RAs was normalized as 59 (72*1000/1222) and 49 (49*1000/1061) respectively. The tables provided in the following section where the move cycles found in both subcorpora are discussed include the normalized values for the frequency of occurrences. These tables also show the number of RAs in which the patterns are observed in both quantitative and qualitative subcorpus.

4.1.2.1 M2-M3 & M3-M2 Cycles

The most frequent move cycle observed in both types of RAs was **M2-M3** (*reporting selected results–discussing the results*) that occurred 59 times in 18 (100%) of the quantitative and 49 times in 12 (67%) of the qualitative RAs. In 8 of the qualitative RAs, the authors switched back to M2 after an M2-M3 sequence **(M2-M3-M2)**. In these qualitative articles, they first reported what was stated by their participants (M2S2), then interpreted what these statements suggested (M3S1); and then reported further statements provided by them to support or exemplify their interpretations (M2S2) (i.e. **M2S2-M3S1-M2S2**). Table 14 shows that the authors achieved **M2-M3** pattern with the steps of **M2S2-M3S1**, *reporting result(s) with statistics and/or examples* and then *interpreting*, 20 times in 12 quantitative RAs and 33 times in 11 qualitative RAs. In this cycle, the authors mainly used the first substep of M3S1 (M3S1A) through which they *made claims based on their findings* (**M2S2-M3S1A**).

Table 14. M2-M3 Cycles Observed in Corpus Data

Patterns	Quan		Qual	
	f (times)	N of RAs	f (times)	N of RAs
M2-M3 (reporting-discussing)	59	18	49	12
M2S2-M3S1 (rep. res. with stat./exmp.-interpreting results)	20	12	33	11
M2S2-M3S1A (reporting results with stat./exmp.-interpreting with claims)	18	11	29	11
M2S2-M3S1-M3S2 (reporting- interpreting results-explaining reasons for results)	4	4	12	7
M2S2-M3S3 (reporting results with stat./exmp. Comparing results)	18	10	-	-
M2S2-M3S3A (reporting results with stat./exmp. stating if results support other hyp./res)	15	9	-	-
M2-M3-M2	-	-	15	8
M2S2-M3S1-M2S2	6	6	12	7

In 7 of the qualitative RAs, the pattern of *reporting-interpreting* (M2S2-M3S1) was followed by *explaining the reasons for results* (M3S2). This chain of **M2S2-M3S1-M3S2** was observed 12 times in these articles and emerged as a pattern only in qualitative data.

Table 14 also reveals the steps of M2-M3 that formed a pattern only in the quantitative RAs. The sequence of **M2S2-M3S3**, i.e. *reporting result(s)* and *comparing (them) to previous finding(s)* was observed 18 times in 10 RAs. In 15 of these cases, after reporting a certain finding, the authors stated if that finding is congruent with the findings or claims of other authors **(M2S2-M3S3A)**.

The move cycle analysis through the n-gram facility of AntConc identified the sequence of **M2S1-M3S1** through which the authors first *stated if their result(s) confirmed the hypotheses* of the study or not and then *interpreted* them 6 times (normalized value = 5) in 4 quantitative RAs. This cycle was not included in the Table 14 as its normalized frequency of occurrence is below the cut off point of 12 used as a criterion. However, if it is considered that M2S1 was identified only 12 times in the whole quantitative data (see Table 7), in half of its cases it was followed by authors' interpretation and therefore it is worth mentioning as a possible pattern. Furthermore, this pattern was not observed at all in the qualitative subcorpus, because M2S1 was identified as a move that is specific to the quantitative RAs only.

The sequence of **M3-M2** (*discussing the results- reporting selected results*) was the third most frequently observed pattern in the corpus data. It was observed 32 and 37 times in 13 of the quantitative and 12 of the qualitative RAs respectively. As shown in Table 15, it was mostly achieved by the steps of **M3S1A-M2S2** (*interpreting what is suggested by the findings and making claims/arguments based on them-reporting results with statistics and examples*). In the qualitative subcorpus, M3-M2 was followed by M3 (**M3-M2-M3**) 34 times in 12 RAs; by M3-M2 (**M3-M2-M3-M2**) 14 times in 8 RAs, and by M3-M2-M3 (**M3-M2-M3-M2-M3**) 12 times 7 RAs. These numbers of occurrences reveal that the authors of qualitative RAs used *discussing-reporting-discussing-reporting* cycles more often than the authors of the quantitative RAs. Besides, the cycle of M3-M2-M3 achieved by the steps of M3S1A-M2S2-M3S1A (*interpreting what is suggested by results and making claims based on them-reporting results with relevant statistics and examples- interpreting what is suggested by results)* is a pattern found only in qualitative RAs. In other words, the authors of qualitative research stated what is reported by their participants, interpreted what those statements suggested, and provided other examples of their statements that they interpreted

again.   In the quantitative data where **M3S1A-M2S2-M3S1A** was not observed at all, however, the authors might have achieved the pattern of **M3-M2-M3** using a combination of other M3 steps like M3S2 (explaining reasons for results) and M3S3 (comparing results to those of other studies) along with M3S1 (interpretation of results).

Table 15. M3-M2 Cycles Observed in Corpus Data

Patterns	Quan		Qual	
	f (times)	N of RAs	f (times)	N of RAs
M3 M2 (discussing-reporting)	32	13	37	12
M3S1-M2S2 (interpreting results- reporting results with stat./exmp.)	11	8	22	10
M3S1A-M2S2 (interpr. with claims - reporting results with stat./exmp.)	11	7	22	10
M3-M2-M3	25	12	34	12
M3S1-M2S2-M3S1	-	-	16	7
M3S1A-M2S2-M3S1A	-	-	14	8
M3-M2-M3-M2	7	7	14	8
M3-M2-M3-M2-M3	5	5	12	7

4.1.2.2 M3-M1 & M1-M3 Cycles

M3-M1 (*discussing results-setting background)* was the second most commonly used cycle by the authors of the corpus in this study. Within the context of M3, the authors referred to the background information, especially conceptual background, using the step M1S2. In other words, while *interpreting the results* of their study (M3S1) they sometimes

needed to remind *general conceptual background* (M1S2) that set the context for their results. The Table 16 shows how many times the authors of the quantitative and qualitative RAs were engaged in this cycle of M3S1-M1S2. As revealed by the numbers, the authors of the quantitative RAs used the cycles of **M3-M1** and **M1-M3** more often than the authors of the qualitative RAs. However, the only steps of M3 and M1 that formed a pattern, namely M3S1-M1S2 and M1S2-M3S1, were observed more often in the qualitative RAs. What is more, the latter observed 12 times in 7 of the qualitative RAs formed a pattern in this subcorpus only, which is congruent with the previous result of this study that revealed authors' tendency to provide more conceptual background in the discussion sections of the qualitative RAs.

Table 16. M3-M1& M1-M3 Cycles Observed in Corpus Data

Patterns	Quan		Qual	
	f (times)	N of RAs	f (times)	N of RAs
M3-M1 (discussing-backg. info.)	34	17	30	14
M3S1-M1S2 (interpreting results-conceptual backgr.)	9	8	12	9
M3-M1-M3	15	12	10	7
M1-M3 (backg.info.–discussing)	28	15	24	11
M1S2-M3S1 (conceptual backgr.-Interpreting results)	4	4	12	7

4.1.2.3 M1, M2 & M3 cycles

M1-M2 cycle through which the authors first provided background information regarding the study and then reported the main findings was observed 22 times in 14 of the quantitative RAs and 15 times in 8 of the qualitative RAs. Among the steps of M1, providing background in relation to the methodology of the study (M1S1) before reporting its findings

(M2S2) appeared as a pattern in quantitative research reports with 12 times of occurrences in 7 RAs as shown in Table 17. This observation can be explained by the previous finding of the present study that revealed authors' tendency to provide more methodological background information in quantitative RAs then in qualitative RAs. Previously it was found that M1S1 made up 6.38% of all quantitative moves but 3.48% of all qualitative moves.

Table 17. M1, M2 & M3 Cycles Observed in Corpus Data

Patterns	Quan		Qual	
	f (times)	N of RAs	f (times)	N of RAs
M1-M2 (backg.info-reporting res.)	22	14	15	8
M1S1-M2S2 (methdol.backg..- reporting results with stat./exmp.)	12	7	7	3
M1-M2-M3	17	12	9	5
M1-M3-M2	13	9	5	4
M3-M1-M2	12	8	6	3
M3-M1-M2-M3	12	8	-	-

Table 17 also reveals that authors use a combination of M1 (background information), M2 (reporting results), M3 (discussing results) cycles like M1-M2-M3 (background-reporting- discussing), M1-M3-M2 (background-discussing-reporting) and M3-M1-M2 (discussing-background-reporting) more often as a pattern in quantitative RAs. The four-move cycle of M3-M1-M2-M3 (discussing-background-reporting-discussing), on the other hand, was not identified at all in qualitative RAs.

4.1.2.4 *M3-M6 & M6-M3 Cycles*

In most of the RAs within the corpus, the authors used the cycle of M3-M6 through which they *discussed their findings* and then *made implications* (M6) based on them. As shown in Table 18, this cycle (M3-M6) was observed 30 times in 17 quantitative RAs and 35 times in 15 qualitative RAs. Among all the steps of M3 and M6, *interpretation of a result* (M3S1) immediately followed by its *pedagogic implications* (M6S2) was observed 19 times in 12 of the qualitative RAs, which formed a pattern in this subcorpus. This finding reveals authors' tendency to immediately relate an interpretation to some practical classroom implications in qualitative research. However, this rhetorical behavior was less (8 times in 7 RAs) identified in quantitative research. This can be attributed to the observation that in quantitative research articles, the authors used other steps of M3 more often before drawing pedagogic and research implications. For example in 5 of them, the authors used the cycle of M3S3-M6S1 (comparing results to those of previous studies-research implications) 7 times and the cycle of M3S3-M6S2 (comparing results to those of previous studies-classroom implications) 4 times.

Table 18. M3-M6 & M6-M3 Cycles Observed in Corpus Data

Patterns	Quan		Qual	
	f (times)	N of R	f (times)	N of RAs
M3-M6 (discussing-drawing impl.)	30	17	35	15
M3S1-M6S2 (interpreting results-pedagogic impl.)	8	7	19	12
M3S1A-M6S2A	5	4	12	8
M6-M3	15	12	19	12
M6S2-M3S1	6	4	12	7
M6-M3-M6	10	8	12	6

The Table 18 also shows the cycles of M6-M3 and M6-M3-M6 in both subcorpus. The authors who drew an implication (M6) referred to an interpretation of results (M3) and drew another implication (M6) 12 times in 6 of the quantitative and 10 times in 8 of the qualitative RAs respectively.

4.1.2.5 *M5-M6 & M4-M6 Cycles*

The Table 19 shows other moves that co-occurred with the *implications* move of M6 within the RA discussions. According to this table, the authors drew implications based on their results (M6) right after they evaluated the study (M5) 21 times in 13 of the quantitative RAs and 11 times in 6 of the qualitative RAs. In more than half of these M5-M6 sequences (i.e. 12 times in 10 quantitative RAs, and 6 times in 4 qualitative RAs), the authors evaluated the study by acknowledging its limitations (M5S1) before making any methodological suggestions for future studies (M6S1B). In other words, they based methodological suggestions on their own methodological limitations. The cycle of these two steps, M5S1 and M6S1B, which had already been identified as making up more of all quantitative moves than

of qualitative moves in the corpus data (see *Figure 7 & 8)*, emerged as a pattern followed by the authors of the quantitative RAs who also used the cycles of M6-M5 and M6-M5-M6 more often than the authors of the qualitative RAs.

Table 19. M5-M6, M1-M6, M4-M6 Cycles Observed in Corpus Data

Patterns	**Quan**		**Qual**	
	f (times)	N of RAs	f (times)	N of RAs
M5- M6 (evaluating study-draw. impl.)	21	13	11	6
M5S1-M6S1 (ackn. limitations- research implications)	12	11	8	5
M5S1-M6S1B (acknowledging lim.- methodological suggestions for research)	12	10	6	4
M6- M5	17	12	10	6
M6-M5-M6	12	10	7	5
M4- M6 (summarizing the study- drawing implications)	14	11	10	10
M4S2-M6S2 (summarizing findings- drawing pedagogic implications)	12	9	7	7
M6- M4 (drawing implications- summarizing the study)	10	8	13	10

The Table 19 also displays that the authors presented their implications within the context of M4, *summarizing the study,* in both types of RAs. In the quantitative RAs, they restated the profound findings (M4S2) the pedagogic implications of which they discussed (M6S2). This sequence of M4S2-M6S2 observed 12 times in 9 of the quantitative RAs was identified as a pattern followed in such research articles. In the qualitative RAs, however, the

sequence of M6-M4 *(implications-summary of the study)* appeared as a pattern with 13 occurrences in 10 of the articles.

4.1.2.6 Summary of the Move Patterns

The findings of the analysis conducted to find move patterns within the corpus data confirmed that the moves and steps identified in RA discussions tend to occur in cycles as stated in previous move studies (including Swales, 1990; Holmes, 1997; Posteguillo 1999; Lewin, Fine and Young, 2001; Peacock, 2002). The authors did not follow a fixed order through *move 1* to *move 6*; instead, depending on the communicative purpose they aimed to achieve, they used a variety of move/step sequences some of which emerged as patterns. These sequences observed as patterns in the corpus of the study are summarized in the following Table 20. According to that, at move level, both types of RAs included the patterns of **M2-M3** *(reporting results- discussing results)* and **M3-M2**; **M3-M2-M3** *(discussing - reporting-discussing)*; **M3-M1** *(discussing-providing background information)* and **M1-M3**; **M1-M2** *(providing background information-reporting results)*; and **M3-M6** (*discussing results-drawing implications*) and **M6-M3**. In other words, at both types of RAs, **M3** mostly co-occurred with **M2, M1** and **M6**.

However, the cycles of **M3-M1-M3** *(discussing results-providing background- discussing results);* **M1-M2-M3** *(providing background- reporting results-discussing results)¸* **M3-M1-M2** and **M3-M1-M2-M3** *(discussing results- providing background- reporting results-discussing results);* **M5- M6** *(evaluating study-drawing implications),*

Table 20. The Summary of the Patterns Identified in Both Types of RAs

Quan Patterns	Qual Patterns
M2-M3 (reporting- discussing)	M2-M3
M2S2-M3S1(A)	M2S2-M3S1(A)
--	M2S2-M3S1-M3S2
M2S2-M3S3(A)	NO*
NO*	M2-M3-M2
--	M2S2-M3S1-M2S2
M3-M2	M3-M2
	M3S1(A)-M2S2
M3-M2-M3	M3-M2-M3
NO*	M3S1-M2S2-M3S1
NO*	M3S1A-M2S2-M3S1A
--	M3-M2-M3-M2
--	M3-M2-M3-M2-M3
M3-M1 (discussing-background info.)	M3-M1
-	M3S1-M1S2
M3-M1-M3	--
M1-M3	M1-M3
--	M1S2-M3S1
M1-M2 (background-reporting results)	M1-M2
M1S1-M2S2	--
M1-M2-M3 (backg.-reporting-discussing) --	
M3-M1-M2	--
M3-M1-M2-M3	NO*
M3-M6 (discussing-implications)	M3-M6
M3S1(A)-M6S2(A)	-
M6-M3	M6-M3
-	M6S2-M3S1
-	M6-M3-M6
M5- M6 (evaluating study-implications)	-
M5S1-M6S1(B)	-
M6-M5	-
M6-M5-M6	-
M4-M6 (sum. study-implications)	-
M4S2-M6S2	-
-	M6-M4

* Not observed at all

M6-M5 and **M6-M5-M6** (*drawing implications-evaluating study-drawing implications*); and **M4-M6** *(summarizing the study-drawing implications)* made patterns only in quantitative RAs. Among these sequences, the co-occurrence of the *evaluation* move of **M5** and the *implication* move of **M6** stands as a pattern only within the quantitative RAs.

On the other hand, some cycles observed in both types of RAs made patterns only in the qualitative RAs: **M2-M3-M2** *(reporting results-discussing results-reporting results)*, **M3-M2-M3-M2** and **M3-M2-M3-M2-M3; M6-M3-M6** (*drawing implications-discussing results-drawing implications*); and **M6-M4** (*drawing implications-summary of the study).* However, the move cycles of **M2-M3-M2** and the sequences that included the steps of **M3S1A-M2S2-M3S1A** through which the authors *interpreted the results-reported the results- interpreted the results* were not observed at all in the quantitative RAs and appeared as patterns peculiar to the qualitative RAs only. Similarly, the move sequence of **M3-M1-M2-M3** and the sequences with steps of **M2S2-M3S3A** (*reporting results with statistics – comparing findings to those of other studies)* emerged as patterns of quantitative RA discussions.

4.2   Findings of the lexico-grammatical analysis

This section reports and discusses the results of the lexico-grammatical analysis to answer the research question s 2.1 and 2.2, which aim to identify the recurrent patterns, namely 4- and 5-part of speech sequences (PoS-tag bundles) and their lexical expressions, unusually frequent in a certain discussion move in reference with their occurrence in the other moves of the study. It should be noted that since the present study does not use an independent reference corpus to compare the bundles identified, it abstains from adopting the term 'lexical bundles'. Instead, it uses the term formulaic expressions and structures to refer to the lexical combinations that could only be identified by their PoS categories due to the

small size of corpus data in each move (see Appendix L). In other words, what was found recurrent in the moves was not necessarily the lexical expressions but their PoS categories. The keyness of these recurrent PoS categories in a certain move was determined by the log likelihood statistical values as mentioned earlier in the methodology section of this study. It should also be noted that formulaic PoS bundles the keyness of which was shown for a certain move were observed in other moves as well, but their occurrence in those moves were not found unusually frequent.

The examples of these expressions identified as key PoS-tag bundles in each move are presented in the following sections with the discussion of their structural and functional features. The tables provide the abbreviations used in Penn Tag List to refer to the PoS tags. The full list of these bundles, their frequency (tokens), expected, and log likelihood values, and the codified corpus data in which they were observed are provided in Appendices M-R.

4.2.1   Five-PoS tag and 4-PoS tag bundles in Move 1

As shown in Table 21, the study identified one 5-PoS tag, and two 4-PoS tag categories that are unusually frequent in *Move 1- Setting Background* which comprised a 10,225-word subcorpus (including units analyzed from both quantitative and qualitative RAs) within the whole corpus (See Appendix L).

Structurally, the 5-PoS tag bundles included a verb with past tense, and a noun phrase with a prepositional phrase fragment (*of, between,* and *in*) that has been realized 8 times in the first step of Move 1, which provides methodological information about the study (See Appendix M). These bundles revealed lexical expressions like *question examined the relationship between, research examined the usefulness of, researcher addressed the role of, study examined the relevance of,* and *study investigated the relationship between.* Functionally, the authors announced the purpose and the research questions of their studies to

the readers with these expressions. The expression of *None discussed the procedure of ,* however, came from the data identified with the function of M1S2A which set conceptual background by referring to the established knowledge. With this particular expression, the author pointed to a gap in the literature to justify his/her reason to do the current study.

Table 21. 5-PoS and 4-PoS tag Bundles in M1

PoS Bundles		Type	Examples
Verb-based	5-tag: nn vbd dt nn in Noun+ active verb, past tense + noun phrase with prepositional phrase fragment	8	M1S1 *…question examined the relationship between…* *…research examined the usefulness of…* *…researcher addressed the role of…* *…study examined the relevance of…*
		1	M1S2 *None discussed the procedure of…*
	Sum	9	
Verb-based	4-tag: nn vbd dt nn Noun+active verb, past tense + noun phrase	10	M1S1 *… question concerned the extent…* *… study investigated the effects…*
		1	M1S2 *…input was a strategy…*
	rb vbn in jj Adverbs + active/passive verb + prep/that- clause fragment	1	M1S1 *… often given in intensive…*
		6	M1S2 *… predominantly focused on oral…* *… well documented that second…* *… widely argued that formulaic…*
	Sum	18	

Similarly, 4-PoS tag bundles that were found to be typical of Move 1, produced a total of 18 types of lexical structures that reflected the communicative purpose of this move. Eleven of them involved structures with noun and verb phrases with active voice in past tense, some of which had already been captured by the 5-PoS tag category mentioned above (See Appendix M); and were generally uttered to provide methodological background information

(M1S1). Seven of them, however, included adverbs (time adverbs *often* and *usually,* additive adverb *also,* manner adverb *well,* degree adverb *widely* and *predominantly;* and stance adverb *truly*) and prepositional or that-clause fragments following active (e.g. have *also shown that implicit*) and passive verbs (is *often given in intensive*) in present tense. These formulaic structures were generally uttered to set conceptual background for the study (M1S2A). It should be noted that 4 of the adverbs used in these formulas, namely *also, often, usually* and *well,* were listed by Biber, Johansson, Leech, Conrad, and Finegan (1999) among the most frequent adverbs that occur at least 200 times per million words in academic prose. The findings suggested that the authors of the corpus data in this study used these adverbs when they referred to the established knowledge to set the theoretical background.

4.2.2 Five-PoS tag and 4-PoS tag bundles in Move 2

The study identified 37 different types of lexical structures realized by five 5-PoS tag bundles in 8,150- word Move 2 through which the authors reported the selected results with statistics and examples (M2S2), or in an integrated way (M2S3). These PoS bundles that emerged as typical of M2 were mainly verb and prepositional-based as shown in the following Table 22 (See full list in Appendix N).

Among the 13 different types of prepositional phrases, *in the present study* followed by a comma *(,)* that occurred 6 times in 5 different texts appeared as the most frequently used prepositional bundle. One plausible explanation for its frequent use may be authors' efforts to relate the newly found information to the present study of their own. This bundle was followed by two prepositional phrases that semantically functioned as linking adverbials that bind the units of discourse by marking a contrast (*on the other hand,*) and an addition (*at the same time,).* In Move 2, they were identified twice in two separate texts (See Appendix N). It should be noted that *in the present study, on the other hand* and *at the same time* had already been identified as lexical bundles in the previous studies (Qin, 2014; Cortes, 2013; Cortes,

2004). In fact, Cortes found the exclusive use of *on the other hand* to make topic generalizations while establishing the territory (Move 1, step 2) and the use of *in the present study* to announce present research descriptively while occupying the niche (Move 3, steps 1) in the introduction sections of RAs. Therefore, they can be referred to as formulaic lexical bundles in this study as well, unlike the other lexical structures listed in the following table that may or may not be bundles at all. Yet, their PoS categories including a prepositional phrase fragment emerged as a PoS bundle in the present study.

Table 22. 5-PoS tag Bundles in M2

PoS Bundles		Type	Examples
Preposition-based	in dt jj nn, Prepositional phrase + (,)	13	M2S2 *At the same time,…* *…for the traditional treatment,…* *In the present study,…* *…of a statistical relationship,* *On the other hand,…*
	Sum	13	
Verb-based	vbd vbn to vb jj Passive verb, past tense + to-clause fragment	4	M2S2 *…was found to be useful…* *…was theorized to be integral…* *…were felt to have different…* *… were made to include relevant*
		1	M2S3 *… were found to be significant…*
	nn vbd in dt nn Noun phr. +active verb, past tense + prepositional /that - clause fragment	5	M2S2 *…analysis revealed that the number…* *…informant commented on the phrase* *…result revealed that all task…* *… t-test revealed that the study …*
	vbd dt jj nn in Active verb, past tense +adj. phr.with prep.phrase fragment	6	M2S2 *…confirmed the textual analysis in…* *… invented a high number of…* *… revealed a significant variation in…*
		2	M2S3 *… had a positive effect on…* *…had a positive impact on…*

	nns vbd dt nn in Noun phr.+ active verb, past tense + noun phr. with prep.phrase fragment	6	M2S2 *… participants expressed an interest in…* *… students mentioned the importance of…* *… students reported an interest in…* *… teachers cited a range of…*
	Sum	24	

Other key PoS bundles in Move 2 involved verb phrases preceded by nouns (*participants expressed an interest in);* and followed by noun phrases with prepositional phrase fragments (*teachers cited a range of, writers boosted the statement in),* by adjective phrases (*confirmed the textual analysis in, identified a significant relationship between)*, by that-clauses (*analysis revealed that the number,* t-*test revealed that the study*) and by to-clauses having adjectives (*was found to be useful, were found to be significant).*

As for the 4-PoS tag bundles in Move 2, the authors used 209 different types of formulaic structures identified by13 PoS categories to report their findings. 187 (89%) of them were found in units identified with the function of the second step of M2 (See Appendix N). Of all lexical structures, 82 (39.2 %) was made up of prepositional phrases; and 5 (2.4%) of noun-phrases that included quantitative expressions involving cardinal numbers and percentages as in *40 % of those, 15 % of the, 5.5 % for the, one third of the,* and *8 hr on a.* The identification of such expressions as typical of Move 2 was expected as they serve to the function of this particular move that aims to present results with statistical and numerical values.

As shown in Table 23, 122 (58.4 %) of 209 types included a verb phrase complemented by that- clauses (*further claimed that the*, *confirmed that the study, commented that many Statistics),* to-clauses (*was found to be*), adjective phrases (*drew particular interest from, was also implicit, identified a significant relationship*), and noun phrases (*emphasized the necessity for*). In some cases, however, they were preceded by adverbs (*simply meant that*

138

a), adjective phrases (*metalinguistic activity was about)* and part of noun phrases (*informant commented on the*).

Table 23. 4-PoS tag Bundles in M2

PoS Bundles		Type	Examples
noun -based	cd nn in dt	5	M2S2 …*40 % of those…* …*8 hr on a…* …*5.5 % for the…*
	Sum	5	
Verb-based	rb vbd in dt Adverbial phrase + prepositional/that-clause fragment	7	M2S2 …*further claimed that the…* …*further showed that both…* …*simply meant that a…* …*quickly waned in the…*
	vbd in dt nn Active verb, past tense+ prepositional/ that-clause fragment	23	M2S2 …*asked for an example…* …*commented on the phrase…* …*confirmed that the study…* …*found that the students…* …*indicated that the students…* …*occurred across a range…* …*revealed that the study…*
	vbd vbn to vb Passive verb, past tense + to-clause	7	M2S2 …*was found to be …* …*was required to become…* …*were found to account…* …*were found to engage..*
	vbd rb jj in be, past tense + adj. phrase	5	M2S2 …*was also implicit.* …*was not significant at…*
		1	M2S3 …*were comparatively few of…*
	vbd in jj nns Active verb + prepositional phrase/ that-clause fragment	7	M2S2 …*commented that many Statistics…* …*emphasised that authentic data…* …*mentioned that numerous conversations…* …*occurred in different forms…*
	vbd jj nn in Verb and adjective phrase with prep.phrase fragment	3	M2S2 …*drew particular interest from…* …*made significant improvement on…* M2S3 …*expressed considerable enthusiasm for…*
		3	…*showed significant improvement for…*

	vbd dt jj nn Verb and adj.phrase	14	M2S2 *...confirmed the textual analysis...* *...followed a native speaker...* *...found the corpus data...* *...identified a significant relationship...* *...invented a high number...*
		2	M2S3 *...had a positive impact* *...had a positive effect...*
	jj nn vbd rb Adjective phrases + be, past tense + adv.	6	M2S2 *...metalinguistic activity was about* *...statistical relationship were relatively...* *...traditional treatment was not...*
		1	*M2S3* *...final output was not*
	vbd dt nn in Active verb, past tense+ noun phrase with prep.phrase fragment	24	M2S2 *...emphasized the necessity for...* *...illustrated a predominance of...* *...mentioned the importance of...* *...provided the correction in...* *...reported an average of...*
	nn vbd jj nn Noun phr.+active verb, past tense + adj. phr.	8	M2S2 *...apprentices saw linguistic competence...* *...participants identified sociocultural differences...* *... types had differential effects...*
		1	*M2S3* *...learners showed significant improvement...*
	nn vbd in dt Noun phr. + verb, past tense+ prepositional/ that-clause fragment	9	M2S2 *...Analysis demonstrated that the...* *...informant commented on the...* *...staff agreed with this...* *...questionnaire indicated that the...* M2S3 *...informant felt that the...*
		1	
	Sum	122	

Prepositional -based	in dt jj nn Prepositional Phrase fragment	4	M2S1 *… by the insignificant difference…* *… in the experimental group…*
			M2S2
		72	*… as an analytical tool…* *… at the usual levels.* *… for the traditional treatment.* *…through a bare assertion.*
			M2S3
		6	*… for the apparent superiority* *… in the second administration*
	Sum	82	

It should also be mentioned that the authors used past tense to report the selected findings of their studies. The formulaic Move 2 PoS bundles involved the use of 53 different verbs in active voice. The most frequently used ones included *was/were* (20 times), *had* (11 times), *showed* (5 times), *said* (4 times), *revealed* (4 times), *found* (3times), *identified* (3), *made* (3 times), *suggested* (3 times), and *meant* (3times).

4.2.3 Five-PoS tag and 4-PoS tag bundles in Move 3

Since Move 3 through which the authors discussed the findings of their studies by interpreting what is suggested by them and by accounting for them has been the largest of all moves in the corpus data with its 22,626-word size, the expectation was to identify the largest group of lexical expressions as typical of this move. However, all of the 5-PoS tag bundles and almost all of the 4-PoS tag bundles that have the highest log likelihood values revealed citations (356 and 431 tokens) to be the most typical of Move 3 as shown in Appendix O. This finding coincides with the previous findings of the present study that unveiled authors' frequent tendency to refer to the works of other authors for different purposes while discussing their results. These purposes had been classified as different subcategories of the steps in Move 3, namely M3S1B (interpreting results by referring to established knowledge on topic), M3S2B (explaining reasons for results by referring to other authors' explanations), M3S2C (making evaluative comments about others' explanations) and M3S3A, B, C

(comparing with results of previous work-giving information about other studies-evaluating them.).

The same 5-tag categories that yielded citations were also observed in Move 1 (126 tokens) and Move 6 (78 tokens). In other words, as shown by move analysis, the authors frequently referred to the relevant literature while setting conceptual background (M1S2) and drawing implications (M6S1-M6S2).

The only unusually frequent M3 category having no citations with a log likelihood above the threshold value has been a 4-PoS tag category as shown in the following Table 24. That category produced 9 different types of verb-based formulaic expressions which involved the use of *Verb be/have +Verb, past participle* especially in sentences where the authors speculated on the possible reasons for their results (M3S2). This possibility was reflected by the modals *can, may, might,* and *could* that preceded the identified bundles in Move 3(See Appendix O).

Table 24. 4-PoS tag Bundles in M3

PoS Bundles		Type	Examples
Verb-based	vb vbn to dt Vbe/have + verb, past participle+to	2	M3S1 *… be justified to a…* *…be limited to the…*
		6	M3S2 *…be attributed to the…* *…be attributed to a…* *…be related to the…* *…be related to this…* *…be related to some…* *…have contributed to this…*
		1	M3S3 *…have contributed to the…*
	Sum	9	

On the other hand, the only formulaic structure without a modal, *have been found not to be related to this,* combined both perfect aspect and passive voice by retaining "the time orientation ('the past with present relevance') of the perfect aspect while demoting the agent through use of the passive voice" (Biber, Johansson, Leech, Conrad and Finegan, 1999: 483). This combination of aspect and voice was described by Biber et al. (p. 482) as the only complex combination that is common in academic prose due to the high frequencies of both perfect aspect and passive voice in academic genres.

4.2.4 Five-PoS tag and 4-PoS tag bundles in Move 4

Move 4 that summarizes the study in terms of its methodology (M4S1) and findings (M4S2) has the smallest corpus of all moves with its 3,382-word size. Yet, the study has identified two 5-PoS, and one 4-PoS tag categories specific to this move (See Appendix P). As shown in Table 25, these expressions are composed of an adjective phrase that precedes a lexical verb in present tense or copular be (*The present study is… , The present investigation supports…);* and a passive verb followed by a prepositional phrase. These expressions enabled authors to highlight what was done in the present study as an overall summary.

Table 25. 5-PoS and 4-PoS tag Bundles in M4

PoS Bundles		Type	Examples
Adjective – and verb phrase	5-tag dt jj nn vbz dt	3	M4S1 *The present study is an …* *The present study underlines the…*
	Adjective phrases + active verb, present tense+ determine5	2	M4S2 *The present investigation supports the..*
	Sum	5	
Verb-based	vbd vbn in dt nn	4	M4S1 *… was adapted for this study.* *… was used as a measure…* *… were used under the assumption..*
	Passive verb, past tense+ prep. Phrase	1	M4S2 *… was facilitated with the use…*
	Sum	5	
Adjective – and verb phrase	4-tag dt jj nn vbz	3	M4S1 *The present paper reports…* *The present study underlines…*
	Adjective phrases + active verb, present tense	4	M4S2 *…the experimental procedure is…* *The present investigation supports…* *… the present study has…*
	Sum	7	

4.2.5 Five-PoS tag and 4-PoS tag bundles in Move 5

The function of Move 5 was defined earlier in this study as authors' evaluation of their own studies in terms of their weaknesses (M5S1) and strength s (M5S2), and the justification of their methodology (M5S3). The units identified with these steps of Move 5 included 3,829 words. The lexico-grammatical analyses revealed 20 and 33 different types of formulaic expressions represented by three 5-PoS tag categories, and two 4-PoS tag categories, respectively as shown in the following Tables 26 and 27 (See Appendix Q). The 5-tag formulaic structures included noun-phrases (14 types) like … *limitation of the study is…,…significance of this research lies…*that explicitly expressed authors' intentions; and

144

passive structures with some adverbs (*apparently, probably*) or negation (6 types).   In formulaic expressions with negation, the authors justified their methodology by stating what was not considered in the development of their data collection instrument (e.g. *however, the instrument was <u>not designed with that purpose</u> in mind –RA 1_9),* and what was *not* done not for the sake of obtaining reliable results (e.g. *We ensured that the testing procedure did not bias the intervention students by implementing a listening- test type that was <u>not practiced during the strategy</u> instruction program*).   Similarly, they considered what is not included in their analysis as one of the limitations (e.g. *Second, because of the relatively high error rate of EFL learners, close to one - half of their data in the incongruent condition was <u>not included in the reaction</u> time analysis*)

Table 26. 5-PoS tag Bundles in M5

PoS Bundles		Types	Examples
Noun-based	nn in dt nn vbz Noun phrase with of-phrase fragment	2	M5S1 *… limitation of the study is…*
		2	M5S2 *… significance of this research lies …* *… significance of this study is…*
		1	M2S3 *…flexibility of the tool enables…*
	dt nns in dt nn Noun phrase with prepositional phrase fragment	5	M5S1 *… any conclusions about the nature…* *... the findings from this study…*
		2	M5S2 *... the effects of this approach…* *... the findings of the study…*
		2	M5S3 *... the ethics of the study…* *… the limitations of the study…*
	SUM	14	
Verb-based	rb vbn in dt nn Passive verb/verb, past participle + prepositional phrase/ that- clause fragment	3	M5S1 *… probably guessed that some sort..* *... rather bored with the routine.*
		3	M5S3 *… apparently compensated by the provision…*
	SUM	6	

4-PoS tag bundles identified as specific to Move 5 also included noun phrases (*..the effects of this..)*, and passive structures (…is *removed from the context* of…). Unlike 5-PoS tag bundles, however, they included verb-based structures that functioned as adverbials (*As illustrated through the findings...,...* however, *limited in the scale* and *…)* or as verb-based structures that post-modified nouns (…findings *reported in this article*).

Table 27. 4-PoS tag Bundles in M5

PoS Bundles		Types	Examples
Noun-based	dt nns in dt Noun phrase + prepositional phrase fragment	7	M5S1 *... the assessors in this..*
		3	M5S2 *... the effects of this* *...the needs of a ...*
		2	M5S3 *... the implications of the...*
	SUM	12	
Verb-based	vbn in dt nn Adverbial position or post-modifying a noun phrase or passive verb + that- clause fragment	12	M5S1 *… excluded from the analyses...* *… limited in the scale...*
		4	M5S2 *... be seen as a strength...* *... reported in this article...*
		5	M5S3 *... As illustrated through the findings...* *… presented in this article*
	SUM	21	

4.2.6   Five-PoS tag and 4-PoS tag bundles in Move 6

Move 6 that the authors used to make suggestions for future research (M6S1) and classroom teaching (M6S2) is composed of 18,060 words. The move analysis of the present study had identified move 6 earlier as the second most frequently used discussion move in RAs. The lexico-grammatical analyses revealed 11 five- PoS tag categories that produced 123 types of formulaic expressions and 16 four- PoS tag categories that yielded 356 formulaic

expressions that occurred unusually frequent in this move of drawing implications. These PoS categories are provided in the following Tables 28 and 29 with some examples (See full list in Appendix R).

Table 28. 5-PoS tag Bundles in M6

PoS Bundles		Types	Examples
Verb phrases with modals	md vb jj to vb Modals + be + adj. phrase + to-clause	11 4	M6S1 ... *may be necessary to examine*... ... *would be interesting to examine*... ... *would be necessary to establish*... M6S2 … *may be able to draw*... … *may be able to pass*... ...*would be able to predict*...
	md vb vbn to vb Modals + passive verb+to-clause	5 8	M6S1 ... *can be done to promote* ... ...*should be administered to ask* ... ... *should be explored to determine*... M6S2 ...*can be devised to focus*... ...*could be instructed to contact*... ... *should be encouraged to observe*
	md vb dt nn in Modals+ active verb+ noun phrase with prep.phrase fragment	7 8	M6S1 ... *could encompass an examination of* ...*might probe this issue with*.. ... *will be the subject of*... M6S2 … *can increase the amount of*.. … *will foster a sense of*... ... *should develop an awareness of*
	prp md vb jj to Anticipatory it /pers. pronoun +modal+ be+adj. phrase	6 1	M6S1 … *it may be necessary to*... … *it might be worthwhile to*... … *it would be important to*... M6S2 … *They would be able to*
	, prp md vb jj Anticipatory it /pers. pronoun + modal + active verb+ adj. phrase	3 2	M6S1 …, *it might be worthwhile*... …, *it would be interesting*... M6S2 …, *we should provide multiple*... …, *it would be helpful for*...

	nn, nns md vb Noun, plural+ modals +active verb	11	M6S2 … addition, instructors can utilize… … behavior, teachers can promote… … course, departments should provide… … motivation , instructors should attempt… … reading, testers should be…
	nns md vb to vb Noun , plural+ modal+ active verb + to-clause	4 6	M6S1 …researchers may need to examine… … researchers might begin to investigate… … researchers might choose to write… M6S2 … instructors should attempt to incorporate… … materials would seem to provide… … teachers may resort to writing…
	nn md vb dt nn Noun phrase + modal + active verb+noun phrase	5 8	M6S1 … inquiry could encompass an examination… … research might use an eye tracking… … research should explore the beliefs,… M6S2 …approach could include the following:… … connection will benefit all students… … learning will foster a sense…
	nns md vb vbn in Noun plural + modal+passive verb+ prep. phrase	4 7	M6S1 … data should be collected in… …factors should be taken into… … teachers can be recruited from… M6S2 … concerns may be addressed by.. … decisions can be filtered through… … Learners should be provided with…
	rb, jj nns md Adverbial phrase, + adjective phrase + modal	4 2	M6S1 Furthermore, future studies should… Therefore, future studies should… Therefore, vocabulary researchers should… M6S2 Moreover, social activities could… Therefore, English teachers should…
	SUM	106	

	vb dt nn in dt	8	M6S1 ... *examine the effectiveness of these*... ... *determine the applicability of the*... ... *determine the nature of the*..
Verb-phrase without modals	Verb, base +prep. phrase	9	M6S2 ... *have a background in the*... ... *dictate the content of the*... ... *discuss the effectiveness of those*...
	SUM	17	

Table 29. 4-PoS tag Bundles in M6

PoS Bundles		Types	Examples
Verb-phrases with modals	nn nns md vb adj. phrase+ modal + active/passive verbs	7	M6S1 … *point learners may become*... … *morphology measures would be*... … *L2 learners will need* …
		11	M6S2 … *L2 students can be*... … *corpus tools can hel*p… ... *comprehension questions should be*...
	jj nns md vb adj. phrases +modal+ verb, base form	15	M6S1 … *Additional studies will help*... ... *future studies should examine* … ... *vocabulary researchers should use*...
		21	M6S2 … age-*appropriate narratives may be* ... *English teachers might seek*... … *individual instructors can do*...
	jj nn md vb Adjective phrase + modal + verb, base form	19	M6S1 ... *future research should examine*... ... *such research can contribute*... ... *ethnographic study might help*...
		15	M6S2 ... *professional development might benefit*... ... *appropriate balance can be*... ... *reflective discourse would enable*...
	nn md vb nns nouns + modal+ active verb+ noun phrase	6	M6S1 … *context could include teachers '*... ... *research will allow researchers*... … *research will allow researchers*
		6	M6S2 ... *institution should arrange workshops*... ... *research would enable teachers*... ... *team could provide NETs*...

	nns md vb jj nouns+modal+ active verb + adjective phrase	7 16	**M6S1** *… interactions would reveal different…* *… differences will be important…* *… observations might provide important…* **M6S2** *… departments should provide multiple…* *… instructors can utilize specific…* *… universities would permit such…*
	nn md vb dt Noun+modal+verb, base form+ determiner	9 13	**M6S1** *… input may require a…* *… work might increase the…* *… reference may be the…* **M6S2** *… approach may demand a…* *… connection will benefit all…* *… step might provide a*
	nns md vb vbn noun+ modal+ passive verb	11 19	**M6S1** *…activities might be increased…* *… tasks should be explored…* *… variables should be kept..* **M6S2** *… activities could be introduced…* *… concerns may be addressed…* *… goals could be addressed…*
	nns md vb dt noun+modal+active verb	8 9	**M6S1** *… studies should compare the…* *… studies will find a…* *… tasks would be another…* **M6S2** *… authorities will wish the…* *… teachers should adopt these…* *… universities would have an…*
	md vb to vb Modal+ verbs + to- clause	7 14	**M6S1** *… may need to examine…* *… will need to be carefully…* *… would need to be* *… must continue to learn…* **M6S2** *… can elect to focus…* *… may need to be…* *… should strive to provide…*
	md vb dt nn Modal + active verb + noun phrase	15 20	**M6S1** *… can affect the usefulness..* *… might increase the learners…* *… should compare the roles…* **M6S2** *… can help the students…* *… should create a classroom…*

			… would represent a learning
	vb vbn to vb be + V3+to-clause	6 10	M6S1 *… be adapted to optimize…* *… be administered to ask…* *… be explored to determine…* M6S2 *… be asked to consider..* *… be devised to focus…* *… be instructed to contact…*
	md vb jj to Modal + be+adj. + to clause	10 4	M6S1 *…may be necessary to…* *… will be important to…* *… would be interesting to…* M6S2 *… may be able to* *… would be unable to*
	vb nns to vb Verb+ plural nouns+ to-verb	4 10	M6S1 *…allow listeners to compare…* *…help researchers to garner…* *…inform efforts to provide…* M6S2 *…inspire publishers to produce…* *…prepare students to deal…* *…train students to become…*
	md vb vbn in Modal+passive verb+ prep. phrase/that-clause	13 27	M6S1 *…can be explored against..* *… could be investigated under…* *… could be tested in…* M6S2 *… can be argued that…* *… can be suggested that…* *… may be combined with…*
	md vb jj nns Modal + verb, based form, adj. phrase	3 8	M6S1 *… would provide valuable insights…* *… might yield different results…* *… would reveal different stances…* M6S2 *… can make authentic texts…* *… should use individual criteria…* *… would be effective means.*
	SUM	343	
Verb-based	to vb nns pos	1 12	M6S1 *… to check participants '..* M6S2 *… to discern readers '...* *… to enhance students '...* *… to maximize learners '...*
	SUM	13	

The findings revealed that 91 % (106 out of 123 types) of the5-tag lexical expressions and 96 % (343 out of 356) of the 4-tag ones involved a modal, i.e. *can, could, may, might, will, would, should,* and *must.* The only modal that was not observed in these formulaic structures was *shall,* which had also been the least frequently used modal in LSWE Corpus (Biber et al, 1999). This finding suggests that authors show a strong preference for using modals while drawing implications based on the findings of their studies. This is mainly due to the fact that modality reveals a writer's stance by "expressing the degree of (un)certainity of the proposition, or meanings such as permission, obligation, or necessity" as put by Biber et al (1999: 457). Modal auxiliary verbs *can, could, may,* and *might* express permission, possibility and ability; *must,* and *should* mark obligation and necessity; and *will, would* and *shall* express volition/ prediction (p.485). When the 4- and 5- PoS tag lexical expressions identified in Move 6 were relisted by the modal auxiliary verb they involved (see Appendix S), expressions of research implications were found to include mostly *would* (41 times), *should* (37 times) and *might* (32 times) with the meanings of prediction, obligation and possibility. In other words, authors shared their predictions regarding the contributions of their methodological suggestions (e.g. *verbal interactions would reveal, would be interesting to learn);* what they certainly believe the future research is obliged to consider (*should be administered to ask; studies should compare the);* and other issues to be considered or possible benefits of their suggestions (*might probe this issue with; observations might provide important).*

The lexical expressions of pedagogic implications, on the other hand, mainly involved *should* (63 times), *can* (58 times) and *may* (38times) that encapsulate the meanings of ability/possibility (*can+may*) and obligation. In other words, the findings are reinterpreted by the authors in terms of what could possibly be done (e.g. *can help the students*, *language*

teachers may resort) or should be done (e.g. *administration should find ways*) pedagogically. Hence, teachers, students, departments, test-developers and curriculum-makers are offered alternative ways to benefit from the findings of the studies.

Another typical lexico-grammatical feature observed in 4- and 5-PoS tag bundles of Move 6 was the use of *to-* clauses controlled by adjectives in extraposed positions in which the modal auxiliary verbs followed the non-referential dummy pronoun *it*. Move 6 involved 25 types (32 tokens) of formulaic expressions with this structure as shown in the following Table 30.

Table 30. Formulaic Structures in Move 6 with Extraposed to-clauses Complementing an Adjective

_____

M6S1   it *may be necessary to examine* the   *it may be necessary to* examine
       it *might be hard to examine* all   *It might be hard to* examine
       it *might be worthwhile to analyze* the   *it might be worthwhile to* analyze
       it *would be important to examine* not   *it would be important to* examine
       it *would be interesting to examine* the   *it would be interesting to* understand
                                                   (4 times)
       it *would be interesting to see* if   *, it would be interesting* to   (2 times)
       it *would be interesting to learn* how   it *would be interesting to*   (4 times)
       it *would be important* to   it *would be important to* examine not
       it *would be interesting to understand* why
       it *would be worthwhile to proceed* along   *it would be worthwhile to* proceed
       it might be worthwhile to   it *might be worthwhile to* analyze
        it *would be worthwhile to* proceed along it *might be hard to* examine
       it *may be necessary to* examine

M6S2 *,it would be helpful for*

_____

To- clauses in these formulaic structures complement adjectives, i.e. *interesting, worthwhile, important, necessary, hard,* and *helpful,* that "present a stance that is not directly attributed to anyone" (Biber et al. 1999: 723) and act as the logical subject of the sentence due to the presence of dummy *it*. Such *to*-clauses controlled by adjectives had been shown to be quite common in academic prose by Biber and his colleagues. Despite its common use, this structure was not observed at all as part of the PoS tag bundles identified in other moves of

the present study. The reason for this may be the small size of the corpus data. Larger corpora may produce bundles with such structures in other moves as well. However, its occurrence in Move 6 can be explained by Biber et al.'s analysis of the semantic domains of the adjectives that tended to control extraposed *to*-clauses. They showed that adjectives in this position mainly express the meanings of possibility, necessity or importance; ease or difficulty; and evaluation, which coincide with the functions of modality. Within the context of *may, might*, and *would,* the adjectives *interesting, helpful* and *worthwhile* convey author's predictive evaluation of the implication to be drawn in *to*-clause; *important* and *necessary* express the possible contributions of what is to be suggested while *hard* reveals author's predictions of difficulty. The use of these lexical adjectives, hence, completes what is conveyed by modality semantically in this particular move.

4.2.7 Summary of the PoS bundles

Overall, the study identified a total of 57 different PoS bundles that produced 199 types of 5-PoS tag and 632 types of 4-PoS tag (excluding citations found in M3) formulaic expressions in the corpus of the study. As seen in the following Table 31, which shows their distribution by the moves, the deduction move (Move 6) included the highest number of key PoS-tag bundles with the highest number of formulaic lexical expressions.

Table 31. Distribution of 5-PoS tag and 4-PoS tag Bundles and Types

Moves		Number of PoS-tag Bundles (PoS categories)	Number of formulaic expressions (types)
Move 1	5-PoS	1	9
	4- PoS	2	18
Move 2	5- PoS	5	37
	4- PoS	13	209
Move 3	5- PoS	citations	--
	4- PoS	citations & 1	9
Move 4	5- PoS	2	10
	4- PoS	1	7
Move 5	5- PoS	3	20
	4- PoS	2	33 types
Move 6	5- PoS	11	123 types
	4- PoS	16	356 types
SUM	5-PoS	22	199
	4-PoS	35	632

An overwhelming majority of these expressions, that is 86 % (172/199) of 5-tag and 84 % (533/632) of 4-tag ones involved a verb-phrase, while 7% (13/199) of 5-tag and 13 % (82/632) of 4-tag ones involved a prepositional phrase. 7% (14/199) of 5-tag and 3% (17/632) of 4-tag expressions, however, were found to involve a noun-phrase.

CHAPTER V

CONCLUSION

5.1   Concluding Remarks Regarding the Rhetorical Structure of RA Discussions

RAs published in English language as the main academic genre of disseminating scientific knowledge has been the focus of this corpus-based genre analysis study with a particular interest in their least explored components, Discussion sections with conclusion. The study first aimed to identify the rhetorical structure of these sections extracted electronically from the reports of 36 empirical studies with a quantitative and qualitative approach to research in the field of Applied Linguistics/English Language Teaching. Drawing on Swales' move analysis for rhetorical investigation, the study used Yang and Allison's (2004) structural model for RA discussions as the initial coding framework to identify authors' communicative purposes to utter each one of 1172 sentences that came from 18 quantitative RAs and 1010 sentences from 18 qualitative RAs. Hence, a total of 2182 sentences comprising the 66272-word corpus of the study were hand-tagged to identify the moves and steps used by the authors to discuss the findings of their studies.

During the initial coding of these sentences, some utterances were found to have functions that were not specified explicitly in Yang and Allisons' model. This caused coding difficulties and led to the reclassification of Yang and Allisons' 7 functional units with 10 steps by the emerging functional subunits. The labor-intense coding spree ended up with the development of a new structural model that included 6 moves, namely *setting background* (M1), *reporting selected results (M2), discussing results (M3), summarizing the study (M4), evaluating the study (M5),* and *drawing implications (M6),* realized by 17 steps and 14 substeps. Although these moves looked similar to those of the initial model, the new one included more categories that classified the communicative purpose of every single utterance, which provided new insights into how discussion moves are realized.

To name the main differences, M1 that provided *background information* in both models was found to have 3 steps with different functions in the present study. While the initial model did not recognize the functional differences between the statements of methodological and of theoretical background, the new model classified them as separate steps on the grounds that the former included information regarding the study at hand, but the latter involved claims and generalizations based on previous research, and resembled the statements of establishing research territory and the niche as identified by Swales in the introduction sections of RAs. In discussion sections, setting theoretical background enabled the contextualization of the findings to be interpreted and the justification of topic centrality by pointing to a gap in the related literature, which was previously observed by Kanoksilapatham (2005) in biochemistry articles. Besides, the study also identified statements that announced what will be presented in the sections that follow. Therefore, these functionally different types of background information were treated as different steps of M1 in the new model.

Second, M2 that *reported selected results* were found to have 3 steps one of which corresponded to a separate move (M3- Summarizing results) in Yang and Allison's model. This step involved the statements of *integrated results* (M2S3) that authors were observed to use occasionally in the present corpus data as a different strategy of sharing results rather than a major move. Similarly, the statements of *whether or not the results support hypotheses* (M2S1) were also classified as a step of M2, because they were observed only in quantitative RAs due to the deductive nature of that approach. The *reporting of numerical findings* (M2S2), on the other hand, corresponded to Yang and Allisons' M2- reporting results. However, its definition was enlarged to classify the reports of what was stated by the participants of the qualitative research as well.

Third, with respect to the most frequently observed *discussion of the results* move, M3, the study confirmed the steps identified by Yang and Allison: Both models included steps through which the authors interpreted their results, accounted for them, compared them to those of other studies, and evaluated them. However, the present study identified 8 subunits (M3S1A, M3S1B, M3S2A, M3S2B, M3S2C, M3S3A, M3S3B, M3S3C) with different functions that contribute to the realization of these steps. These subunits revealed that in discussion sections authors had more reasons to refer to the previous literature than just to compare their results. Within the context of M3, they referred to other authors' views and explanations to interpret (M3S1B) and account for (M3S2B) their results; and they evaluated the explanations offered by others as shown earlier by Baştürkmen (2010). However, what was not shown previously is that authors also reported the findings of other studies which they mentioned for the purpose of comparison and contrast, they interpreted these findings, proposed reasons for them, and evaluated them (M3S3B, M3S3C). In fact, the findings suggested that *setting conceptual background* (M1S2) that made up 12.44 % of all qualitative moves and 8.75 % of all quantitative moves was the most common reason for the authors to refer to literature. Comparison to other studies (M3S3A) was the second most common reason (6.14 % of all moves) in quantitative RAs. However, in qualitative RAs, M1 step of providing theoretical background was followed by making classroom implications in reference with others (M6S2B), interpreting results with other's views (M3S1B), and then by the M3 step of comparing results to those of previously conducted studies (M3S3A). At this point, it can be concluded that one of the main contributions of the present model is the identification of authors' purposes to refer to established knowledge and previous literature in discussion sections.

As for the other moves, M4 that corresponded to Yang and Allison's M5 summarized the study in terms of what was done (M4S1) and what was suggested overall by the findings

(M4S2). Hence, what Yang and Allison meant by "summarizing the main points from the perspective of the study" was defined more precisely. Similarly, the move of evaluation of the study (M5), which is somewhat equal to Yang and Allison's M6, included the steps in the initial model. However, the definitions of the available steps were revised in a way to classify the weaknesses of the study, especially those caused by methodology, by the step of *acknowledging the limitations* (M5S1), whereas the strengths of the study by the step of *justifying the methodology* (M5S3). This latter step that Yang and Allison referred to as the *evaluation of the methodology* had caused some difficulties in coding because the statements of methodological limitations could also be classified as the evaluation of the methodology.

Finally, the study identified the deduction move of *drawing implications* (M6) as the second most frequently used functional unit in discussion sections. The new model clearly presented *research* and *classroom* implications as two different steps of M6. Yang and Allisons' model, on the other hand, included the steps of *making suggestions* (M7S1)*, recommending further research* (M7S2) and *drawing pedagogic implications* (M7S3), which also caused coding difficulties because the statements that recommended further research (M7S1) naturally proposed suggestions to be considered as well (M7S2). The new model, however, clearly distinguished the functions of *stating and justifying the need for future research* and *making methodological suggestions* as substeps of *research implications* (M6S1)*;* and defined any suggestions to be considered by students, teachers, departments, language programs, and curriculum and test developers by the substeps of *pedagogic implications* (M6S2)*.*

In conclusion, the new model proposed for RA discussions in the present study emerges as a comprehensive structural analysis tool that promises an insightful investigation in further genre analysis studies.

Another purpose of the study was to investigate the rhetorical differences between the quantitative and qualitative RA discussions in terms of move-step occurrences and cycles. The study revealed that all of the 18 quantitative RAs involved the steps of interpreting the results with claims and arguments (M3S1A), comparing them to the findings of other studies (M3S3A), and giving methodological background (M1S1) and methodological suggestions for future research (M6S1B) in their discussions; and therefore, they were considered to be obligatory to use. Authors' use of obligatory quantitative steps of M1S1, M3S3A, M6S1B and a conventional M5 step of acknowledging the methodological limitations, M5S1, was found to be significantly related to the type of research. Since these steps made up more of all quantitative functional units than of all qualitative functional units identified, it is not far-fetched to conclude that the authors of the quantitative RAs are more concerned with methodological issues in their discussions.

In the qualitative RA discussions, on the other hand, no obligatory step was found; but the interpretation of the results (M3S1A) and drawing pedagogic implications with/without reference to literature (M6S2) and research implications (M6S1) were observed to be highly conventional to use. Although not statistically significant, the authors of the qualitative RAs used more instances of providing theoretical background (M1S2), reporting selected results (M2S2), interpreting the results (M3S1), giving the summary of what is suggested by the findings (M4S2) and drawing pedagogic implications (M6S2) in their discussions than the authors of the quantitative RAs.

Finally, the findings of the study provided evidence for the cyclic nature of moves confirming the previously reported studies in the literature (Swales, 1990; Holmes, 1997; Posteguillo 1999; Lewin, Fine and Young, 2001; Peacock, 2002). The study identified the move-step patterns that the authors were inclined to use in both types of RA discussions at 3 levels. The most frequent pattern observed in both quantitative and qualitative subcorpus at

160

move level was M2-M3 and M3-M2 through which the results were discussed before and after being reported. Further step level analyses revealed that in most of M2 and M3 sequences, discussion was frequently carried out with the step of interpretations with claims before and after the selected results were reported with examples and statistical values. The interpretations were also followed and preceded by pedagogic implications and the statements of theoretical background. In a nutshell, these findings confirmed the cycles of *interpretation-reporting the results* (and vice versa)*, interpretation-theoretical background information* (and vice versa), and *interpretation-pedagogic implications* (and vice versa) as move-step patterns in RA discussions in Applied Linguistics/ELT.

However, the patterns of drawing implications (mostly research implications) before and after evaluating the study (mostly by acknowledging its limitations); and after summarizing the study were observed only in quantitative RAs. Drawing implications and then making a summary of the study, on the other hand, was observed only in qualitative subcorpus where other qual-only patterns involved longer sequences with M2, M3, and M6 combinations.

5.2   Concluding Remarks Regarding the Lexico-grammatical Analyses of Discussion Moves

After the move structure analysis and the investigation of move-step patterns, the study aimed to explore the lexico-grammatical features of each move using corpus linguistics tools. The scope of this investigation was determined to be the search of formulaicity in terms of the recurrent grammatical structures involving 5-word and 4-word combinations that are typical of each move. For this purpose, the study adopted a frequency-driven approach advocated in the previous studies (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Biber, Conrad & Cortes, 2004; Biber & Barbieri, 2007; Biber, 2009). The bundles of key PoS categories that occurred over a certain frequency threshold were first identified by a *Key-*

bundle Analyzer Program, and then the corresponding lexical combinations were extracted from the PoS-tagged corpus by means of AntConc 3.4.1.w.

The findings provided corpus-based evidence for the existence of unusually frequent PoS bundles in each move. Structurally, the majority of these bundles involved a verb phrase with to-clause and that –clause fragments, prepositional phrase fragments, noun phrase fragments and adjective phrases having prepositional phrase fragments. The rest of the formulaic expressions were composed of prepositional phrases, and noun phrases. A specific finding that needs to be emphasized here is that modality emerged as a key lexico-grammatical feature of drawing implications (M6) move. In other words, formulaic expressions used to make classroom and research implications involved modal auxiliary verbs.

Another finding is that except for an M2- reporting the results- PoS bundle (*noun phrase+active verb, past tense + adjective phrase)* that revealed 8 types of 5-tag and one type of 4-tag expressions with structurally complete units (e.g. *participants identified sociocultural differences),* the rest of the lexical expressions identified in the study were structurally incomplete units. This finding of the study is in line with the most commonly agreed characteristic of lexical bundles (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Biber, Conrad & Cortes, 2004; Biber & Barbieri, 2007).

5.3   Implications of the Study

The present study has significant research and pedagogic implications. First of all, the study developed a systematic procedure for genre analysis that could be adopted in the rhetorical investigation of any written academic discourse. During that procedure, taking sentence as the coding unit of move analysis without disregarding the contexts in which they were uttered enabled the classification of all functional meanings intended by the authors in

RA discussions that would most likely remain unveiled at higher (e.g. paragraph) levels of analysis. Sentence-level coding revealed the contribution of each utterance to the overall meaning established at discourse level, which led to the development of a comprehensive structural model defining a wide range of communicative purposes by the emerging new categories that were not specified in the previous models. Hence, it provided fuller insight into the conventional and optional ways of composing discussions in this particular genre. Therefore, the move analysis steps described in detail in this study may be inspirational and guiding for genre analysts in future studies.

Second, the study used the facilities of *AntConc 3.4.1w,* a corpus analysis tool, very creatively in order to investigate the cycles of move sequences at three (move-step-substep) levels. It is a common practice to use N-gram facility of such tools to extract the lexical bundles with 2- to 5-word combinations in a corpus. In the present study, however, this facility was used innovatively to find the recurrent move cycles after each sentence in the whole corpus was replaced with the codes of their functional categories. This way, the bundles of consecutive functional categories, namely moves and their subcategories were extracted from the corpus of codes as if they were the bundles of word combinations. This method is noteworthy as it promises more accurate and faster identification of move cycles than any manual analyses, and illustrates an innovative way of integrating corpus linguistic methodologies into genre-analysis studies.

Finally, the study presented an alternative approach to the search of formulaicity in a small corpus. Having considered the strong association between structure and function, it showed how move-specific formulaic expressions could be identified by their PoS categories rather than their lexis. Since the subcorpus of each move was small in size, formulaicity was sought not by the repeated lexical combinations but by the typical grammatical structures serving to the main functional purpose of each move, because the meaning is conveyed

163

through structures after all. This way, the study shed some light into how meaning is established in each move of the RA discussions and the structural features of each rhetorical function in these moves.

As for the pedagogic implications, the findings of the study have direct relevance to the teaching of academic writing. As mentioned earlier, academic writing has its own conventions that alert anyone involved, from students to academics, to the need to learn for success in assigned academic tasks and for being accepted as a publishing member of the discourse community. These conventions were shown to be genre- and discipline-bound by numerous studies since Swales introduced moves and steps as functional units of rhetorical organization available to the authors to compose academic prose. A practical pedagogic implication drawn commonly in these studies, including the present one, has been the necessity of teaching genre-bound rhetorical and lexico-grammatical features. In addition to this, evidence from the findings of the present study shows the necessity of teaching move-bound lexico-grammatical features as well along with the other genre-specific features.

To be more specific, corpus based and genre-focused academic writing courses can be designed as part of graduate programs in all disciplines. Those focused on the analysis of RAs could aim to familiarize students with the differing primary communicative purposes of conventional IMRD sections with a particular focus on the language used to convey these purposes.

Regarding the teaching of discussion sections in the field of Applied Linguistics/ ELT, the present study suggests that academic writing instructors should conceptualize their teaching content considering the structural forms and functions of the 6 discussion moves realized by conventional and optional steps and substeps identified in both qualitative and quantitative RAs. In class, after discussing the different characteristics of these two approaches, the students could be provided with corpus-based materials including real

excerpts taken from the published articles illustrating the authentic realization of moves and move cycles in both types of RAs. Using the model offered in this study as a framework, instructors and students could analyze the rhetorical organization of the published RA discussions. Then, students could be encouraged to get involved in the activities that increase their awareness of the frequent structural forms and formulaic expressions peculiar to specific moves, like the use of modality to draw implications. The students could also analyze their own products to compare and contrast them to the published articles to see the use or lack of move-bound structures and expressions. As suggested by Cortes (2013), raising their awareness of the prefabricated expressions and formulaic structures characterizing each move, like those found in this study, could eventually help them produce RAs that would sound more natural in terms of their language and organization. Therefore, the study suggests that the programs in each discipline and the instructors should support the design of such academic writing courses as part of their curriculum.

5.4   Limitations of the Study and Recommendations for Future Research

The study has some limitations that need to be acknowledged. First of all, the specialized corpus of the study was compiled meticulously in consideration with the issue of representativeness which often caused corpus-based approaches to be criticized. As mentioned in the methodology section earlier, all efforts were deployed to compile a representative corpus; yet, it is necessary to bear in mind that no corpus can include all examples of language, but can only represent a partial account of real language as put by Widdowson (2000). Hence, it can reveal a linguistic phenomenon only if the examples of that phenomenon are present in that partial account of language. Therefore, although it is confidently believed that the comprehensive model proposed here captured all communicative purposes intended in RA discussions that composed the corpus of the study, caution is

165

required to generalize the findings unless more research is conducted along this line in order to verify the identified moves-and-steps and their lexico-grammatical features in some other specialized corpora involving the discussion sections of different RAs published in other journals as well in Applied Linguistics/ELT. The present study suggests that the samples of qualitative and quantitative RAs should be included in these corpora to confirm the rhetorical differences found in this study, and the present framework should be used as coding scheme to confirm its descriptive adequacy.

Another limitation of the study is the small size of language representing each move. It is acknowledged that the larger amount of language could have revealed more move-specific lexico-grammatical features. However, the hand-coding of each sentence as part of move-analysis, as done in the present study, would not be feasible in large corpora. At this point, the study has two recommendations that need to be considered: First, future studies using larger corpora can be conducted as research projects involving many researchers who spent long hours of training in move analysis to share the load of sentence coding. Second, a computer program can be developed for automatic analysis of discussion moves. The only known example of such programs is *AntMover* which was basically designed for the analysis of RA introductions based on Swales' CARS model. The hand-coded data of the present study and those of further similar studies will be of great use in the development of such move-analyzer programs.

A final limitation to the study regards the lack of reference corpus to confirm if the move-bound formulaic expressions are lexical bundles or not. Therefore, the study avoids using the term "lexical bundles" as explained earlier. Besides, the study confined the lexico-grammatical investigation only to the search of formulaicity due to the time limitations. More investigation on this line, however, can offer more insights into written academic discourse.

One final remark should be made about the necessity of conducting follow-up studies to investigate the pedagogic uses of these move-bound expressions in the teaching of RA discussions.

REFERENCES

Ädel, A. & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92.

Amnuai, W. & Wannaruk, A. (2013). Investigating move structure of English Applied Linguistics research discussions published in international and Thai journals. *English Language Teaching, 6*(2), 1-13.

Anthony, L. (2011). AntConc (Version 3.2.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

Baker,P. (2006). *Using corpora in discourse analysis.* Bloomsbury Publishing.

Basturkmen, H. (2009). Commenting on results in published research articles and masters dissertations in language teaching. *Journal of English for Academic Purposes, 8*(4), 241-251.

Berkenkotter, C. & Huckin, T.N. (1993). Rethinking genre from a sociocognitive perspective. *Written Communication, 10*(4), 475-509.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In McH. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp.16-36). Amsterdam/Philadelphia: John Benjamins.

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

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

Biber D. (2004). Lexical Bundles in Academic Speech and Writing. *Practical Applications in Language and Computers*, 165-178.

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. (2006). *University Language: A corpus-based study of spoken and written*

registers. Amsterdam: John Benjamins.

Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written

registers. *English for Specific Purposes, 26*(3)*,* 263-286.

Biber, D., Connor, U. & Upton, T.A. (2007). *Discourse on the move: Using corpus*

analysis to describe discourse structure. Amsterdam/Philadelphia: John Benjamins.

Biber, D. (2009). A corpus-driven approach to formulaic language in English.

International Journal of Corpus Linguistics, 14(3), 275-311.

Breiteneder, A. (2009). English as a lingua franca in Europe: An empirical perspective.

World Englishes, 28(2), 256-269.

Bunton, D. (2005). The structure of PhD conclusion chapters. *Journal of English for*

Academic Purposes, 4(3), 207-224.

Chen, Y.-H. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing.

Language Learning and Technology, 14(2), 30-49.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and*

Psychological Measurement, 20(1), 37-46.

Conrad, S. (1999). The importance of corpus-based research for language teachers.

System, 27(1)*,* 1–18.

Conrad, S. (2004). Corpus linguistics, language variation, and language teaching.

In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching.* Amsterdam/

Philadelphia: 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. (2008). A comparative analysis of lexical bundles in academic history

    writing in English and Spanish. *Corpora*, *3*(1), 43-57.

Cortes, V. (2012). *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.

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

Coxhead, A & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and

    grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147.

Dudley-Evans, T. (1994). Genre Analysis: An approach to text analysis for ESP.

    In M. Coulthard (Ed.), *Advances in Written text analysis* (pp. 219-228). Routledge,

    London.

Durrant, P. and Mathews-Aydınlı, J. (2011). A function-first approach to identifying

    formulaic language in academic writing. *English for Specific Purposes, 30*(1), pp. 58–

    72

Flowerdew, J. (2004). The argument for using English specialized corpora to

    understand academic and professional language. In U. Connor & T.A. Upton (Eds),

    Discourse in the professions: Perspectives from corpus, (pp. 11-33). Amsterdam/

    Philadelphia: John Benjamins.

Flowerdew, J. (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, J. & Forest, R.W. (2009). Schematic structures 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). Continuum International Publishing Group.

Flowerdew, J. & Peacock, M. (2001). Issues in English for academic purposes. In J. Flowerdew and M. Peacock (Eds), *Research Perspectives on English for Academic Purposes,* (pp.8-24). Cambridge University Press.

Fraenkel, J. R. & Wallen, N.R. (2000). *How to Design and Evaluate Research in Education.* (4th Ed.) McGraw-Hill.

Fuertes-Olivera, P. A. (2008). Pedagogical Application of Specialized Corpora in ESP Teaching: The case of the UVaSTECorpus. *Scripta Manent, 3*(2), 68-81.

Gavioli, L. (2005). *Exploring Corpora for ESP Learning.* Amsterdam: John Benjamins.

Ghadessy, M., Henry, A. & Roseberry, R. L. (Eds). (2001). *Small Corpus Studies and ELT.* Amsterdam/Philadelphia: John Benjamins.

Granger, S. (2002). A Bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching,* (pp. 3-33). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Greaves, C. & Warren, M. (2010). What can a corpus tell us about multi-word units? In A. O'Keeffe & M. McCarthy (Eds), *The Routledge Handbook of Corpus Linguistics,* (pp. 212-226). London/New York: Routledge Taylor & Francis Group

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

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*(2), 113-121.

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

    University Press.

Hyland, K. (1996). Writing without conviction? Hedging in scientific research articles.

    Applied Linguistics, 17(4), 433-454.

Hyland, K. (2006). *English for Academic Purposes.* London/New York: Routledge.

Hyland, K. & Hamp-Lyons, L. (2002). EAP: Issues and directions. *Journal of English*

    for Academic Purposes, 1(1)*,* 1-12.

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly,*

    30(4), 693-722.

Johanson, S. (2008). Some aspects of the development of corpus linguistics in the 1970s

    and 1980s. In A. Lüdeling & M. Kytö (Eds), *Corpus linguistics. An International*

    Handbook (pp. 33-53)*.* Berlin: Mouton de Gruyter.

Johns, A. M. & Dudley-Evans, T. (1991). English for specific purposes: International in

    scope, specific in purpose. *TESOL Quarterly, 25*(2), 297-314.

Johns, T. (1991). Should you be persuaded- Two samples of Data-Driven learning

    materials. In T. Johns & P. King (Eds), *Classroom Concordancing*. Birmingham:

    ELR.

Jordan, R.R. (2002). The growth of EAP in Britain. *Journal of English for Academic*

    Purposes, 1(1), 69-78.

Kachru, B. B. (1996). English as Lingua Franca. In H. Goebl, P.N. Nelde, Z. Stary, and

    W. Wölck (Eds). *Kontaktlinguistic,* vol.1 (pp.906-13). Berlin/New York: Mouton de

    Gruyter.

Kanoksilapatham, B. (2003). A Corpus-based investigation of scientific research articles:

linking move analysis with multidimensional analysis. Unpublished doctoral dissertation, Georgetown University, Washington, DC.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes, 24*(3), 269-292.

Karabacak, E. (2009). *Unplanned terminology development: A synchronic and diachronic study on economic terms in English newspapers.* Unpublished doctoral dissertation in Applied Linguistics, Northern Arizona University, USA.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.

Kwan, B. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes, 25*(1), 30-55.

Lewin, B.A., Fine, J., & Young, L. (2001). *Expository Discourse: A genre-based approach to social science research texts.* New York: Continuum.

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.

Lüdeling, A.& Kytö, M. (2008). 'Introduction', in A. Lüdeling & M. Kytö (Eds), *Corpus linguistics. An International Handbook.*. Berlin: Mouton de Gruyter, pp. v-xii.

McEnery, T. & Wilson, A. (1996). *Corpus Linguistics: An Introduction.* Edinburgh University Press.

McMillan, J.H. & Schumacher, S. (1989). *Research in Education: A conceptual Introduction (2nd Edition)*. Scott, Foresman and Company.

Mauranen, A. (2005). English as lingua franca: An unknown language? In G. Cortese & A. Duszak (Eds), *Identitiy, community, discourse: English in intercultural settings* (pp.269-93)*.* Bern: Peter Lang.

McEnery, T. & Wilson, A. (1996). *Corpus Linguistics: An Introduction.* Edinburg

University Press.

McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book.* New York: Routledge.

Meyer, H. J. (2004). Global English- a New Lingua Franca or a New Imperial Culture? In A. Gardt, B. R. Hüppauf (Eds), *Globalization and the Future of German* (pp. 65-84). Berlin/New York: Mouton de Gruyter.

Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

Miller, C.R. (1994). Genre as social action. In A. Freedman, P. Medway (Eds) *Genre and the New Rhetoric* (pp.20-36). Taylor & Francis.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes.* Oxford: Oxford University Press.

Nattinger, J. R. & DeCarrico, J.S. (1992). *Lexical Phrases and Language Teaching*. Oxford University Press.

Nesselhauf, N. (1996). 'Learner corpora and their potential for language teaching'. In J. M. Sinclair (Ed), *How to use corpora in language teaching,* (pp.125-151). Amsterdam/Philadelphia: John Benjamins.

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes, 16*(2), 119-138.

Peacock, M. (2002). Communicative moves in the discussion section of research articles. *Systems, 30*(4), 479-497.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System 42,* 220-231.

Scott, M. (2004). WordSmith tools4. Oxfod: Oxford University Press.

Scott, M. & Tribble, C. (2006). Textual Patterns: Key words and corpus analysis in

language education. Amsterdam/Philadelphia: John Benjamins.

Seidlhofer, B. (2004). Research perspectives on teaching English as lingua

franca. *Annual Review of Applied Linguistics, 24,* 209-39.

Seidlhofer, B., A. Breiteneder, & Pitzl, M.L. (2006). English as a Lingua Franca:

Challenges for Applied Linguists. *Annual Review of Applied Linguistics, 26,* 3-34.

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. (2004). Introduction. In J. M. Sinclair (Ed.), *How to Use*

Corpora in Language Teaching, (pp. 1–13). Amsterdam/ Philadelphia: John

Benjamins.

Stemler, Steve (2001). An overview of content analysis. *Practical Assessment,*

Research & Evaluation, 7(17). Retrieved October 24, 2013 from

http://PAREonline.net/getvn.asp?v=7&n=17.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement

approaches to estimating interrater reliability. *Practical Assessment, Research &*

Evaluation, 9(4). Retrieved October 22, 2013 from

http://PAREonline.net/getvn.asp?v=9&n=4.

Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings.*

Cambridge: Cambridge University Press.

Swales, J. M. (2004). *Research genres: Explorations and Applications.* Cambridge:

Cambridge University Press.

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-Rich Part-of-Speech

Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL*, pp. 252-259.

Thompson, G. & Ye, Y. (1991). Evaluating in the reporting verbs used in academic papers. *Applied Linguistics, 12*(4), 365-382.

Thurston, J. & Candlin, C. N. (1998). Concordancing and Teaching of Vocabulary of Academic English. *English for Specific Language, 17*(3), 267-280.

Tribble, C. (2001). Small corpora and teaching writing. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds), *Small Corpus Studies and ELT.* Amsterdam: John Benjamins, pp. 381–406.

Truchot, C. (2002). Key aspects of the use of English in Europe. Strasbourg Language Policy Division. Council of Europe. Retrieved April 14, 2011 from http://www.coe.int/t/dg4/linguistic/source/truchoten.pdf

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

Uysal, H. H. (2012). The critical role of journal selection in scholarly publishing: A search for journal options in language-related research areas and disciplines. *Journal of Language and Linguistic Studies*, *8*(1), 50-95.

Widdowson, H.G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21 (1), 3-25.*

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. (2000). The functions of formulaic language: an integrated

model. *Language and Communication, 20*(1), 1-28.

Wray, A. (2012). What do we (think we) know about formulaic language? An

evaluation of the current state of play. *Annual Review of Applied Linguistics, 32,* 231-

254.

Yang, R. & Allison, D. (2003). Research articles in applied linguistics: moving from

results to conclusions. *English for Specific Purposes, 22*(4), 365-385.

APPENDICES

A.   The List of the Research Articles within the Corpus of the Study

Quantitative Research Articles:

RA1_ Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics, 32*(2), 129-148.

RA2 _ Pichette, F., Serres, de L. & Lafontaine, M. (2012). Sentence Reading and writing for second language vocabulary acquisition. *Applied Linguistics, 33*(1), 66-82.

RA3_ Nassaji, H. (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning, 57*(4), 511-548.

RA4_ Graham, S. & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of French. *Language Learning, 58*(4), 747-783.

RA5_ Potowski, K., Jegersky, J., & Morgan-Short, K. (2009). The effects of instruction on linguistic development in Spanish heritage language speakers. *Language Learning, 59*(3), 537-579.

RA6_ Vandergrift, L. & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning, 60*(2), 470-497.

RA7_ Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning, 60*(3), 534-572.

RA8_ Albert, A. & Kormos, J. (2011). Creative and narrative task performance: An exploratory study. *Language Learning, 61*(1), 73-99.

RA9_Collins, L., Trofimovich, P. & White, J. (2009). Some input on the easy/difficult question: An emprical study. *The Modern Language Journal , 93*(3), 336-353.

RA10_ Hernandez, T. A. (2010). The Relationship Among Motivation, Interaction, and the Development of Second Language Oral Proficiency in a Study-Abroad Context. *The Modern Language Journal, 94(*4), 600-617.

RA11_ Jeon, H. E. (2011). Contribution of morphological awareness to second language reading comprehension. *The Modern Language Journal, 95*(2), 217-235.

RA12_ Huang, S. & Eslami, Z. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal, 96*(4), 544-557.

RA13_ Hwu, F. & Sun, S. (2012). The aptitude-treatment interaction effects on the learning of grammar rules. *System, 40*(4)*,* 505-521.

RA14_ Carreira, J. M. (2012). Motivational orientations and psychological needs in EFL learning among elementary school students in Japan. *System, 40*(2)*,* 191-202.

RA15_ Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*(2), 255-283.

RA16_ Spelman Miller, K., Lindgren, E., & Sullivan, K. P. H. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly, 42*(3), 433-454.

RA17_ Yamashita, J. & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*(4)*,* 647-668.

RA18_ Alptekin, C. & Erçetin, G. (2011). Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL Quarterly, 45*(2), 235-266.

Qualitative Research Articles:

RA19_ Flowerdew, J. & Li, Y. (2007). Language Re-Use among Chinese Apprentice scientists writing for publication. *Applied Linguistics, 28*(3), 440-465.

RA20_ Farrell, T. S. C. & Kun, S.T.K. (2007). Language policy, language teachers' beliefs, and classroom practices. *Applied Linguistics, 29*(3), 381-403.

RA21_ Borg, S. (2009). English language teachers' conceptions of research. *Applied Linguistics, 30*(3), 358-388.

RA22 _ Chang, Y. & Kanno., Y. (2010). NNES doctoral students in English-speaking academe: The nexus between language and discipline. *Applied Linguistics, 31*(5), 671-692.

RA23_ Min, H-T. (2008). Reviewer stances and writer perceptions in EFL peer review training. *English for Specific Purposes, 27*(3), 285-305.

RA24_ Hirano, E. (2009). Research article introductions in English for specific purposes: A comparison between Brazilian Portuguese and English. *English for Specific Purposes, 28*(4), 240-250.

RA25_ Lee, G. (2009). Speaking up: Korean students' oral participation in class discussions in US graduate seminars. *English for Specific Purposes, 28*(3), 142-156.

 RA 26_ Harwood, N., Austin, L. & Macaulay, R.(2010).   Ethics and integrity in proofreading: Findings from an interview-based study. *English for Specific Purposes, 29*(1), 54-67.

RA27_ Testuto, G. (2011). Legal Problem Question Answer Genre across jurisdiction and cultures. *English for Specific Purposes, 30*(4), 298-309.

RA28_ Wharton, S. (2012). Epistemological and interpersonal stance in a data description task: Findings from a discipline-specific learner corpus. *English for Specific Purposes, 31*(4), 261-270.

RA29_ Watzke, J. L (2007). Foreign Language Pedagogical Knowledge: Toward a developmental theory of beginning teacher practices. *The Modern Language Journal, 91*(1), 63-82.

RA30_ Gutierrez, X. (2008). What does metalinguistic activity in learners' interaction during a collaborative L2 writing task look like? *The Modern Language Journal, 92*(4), 519-537.

RA31_ Carless, D. (2007). The suitability of task-based approaches for secondary schools: Perspectives from Hong Kong. *System, 35*(4), 595-608.

RA32_ Luk, J. (2008).Assessing teaching practicum reflections: Distinguishing discourse features of the "high" and "low" grade reports. *System, 36(4)*, 624-641.

RA33_ Koyalan, A. (2009). The evaluation of a self-access centre: A useful addition to class-based teaching? *System, 37*(4), 731-740.

RA34_ Mori, R. (2011). Teacher cognition in corrective feedback in Japan. *System, 39(4),* 451-467.

RA35_ Mcdonough, K. & Chaikitmongkol, W. (2007). Teachers' and learners' reactions to a task-based EFL course in Thailand. *TESOL Quarterly, 41*(1), 107-132

RA36_ Trent, J. (2012). The discursive positioning of teachers: Native-speaking English teachers and educational dicourse in Hong Kong. *TESOL Quarterly, 46*(1), 104-126.

B.   Yang and Allison's (2003) Model for RA Discussions *

Move 1- Background Information

This move relates discussion to the study by recapitulating main points such as research questions, aims and purposes, theoretical or methodological information.

e.g. Our aim has been to explore, within the limits of the data available, a relatively complex issue: the accommodation of languages that parents in ethnolinguistic minority groups have to make …

Move 2- Reporting findings

This is the central Move in which the findings of the study are presented, normally with relevant evidence such as statistics and examples.

e.g. The results indicate that if a subject has a high SR in L1, then it is likely that SR will also be high in L2 …

Move 3- Summarizing results

This Move presents integrated results on the basis of a number of specific results.

e.g. To sum up, it becomes clear that keeping a heritage language alive across generations is not a simple matter of mothers taking a position on language use …

Move 4- Commenting on results

The main purpose is to establish the meaning and significance of the research results in relation to the relevant field.

      Step1 - Interpreting results

      e.g. These results suggest, first, that some significant changes take place between time one and time two and, second, that the knowledge which underlies L2 processing is in some way different to the knowledge which underlies the processing of L1…

*The definitions and examples included here are as given by Yang and Allison (2003: 382-83)

Step 2 - Referring to previous research

e.g. These findings support the previous survey results of Ostler and the ethnographic data of Mason …

Step 3- Accounting for results

e.g. Such differences may also be promoted by the educational systems of both cultures, and by … This can be a reason why …

Step 4- Evaluating results

e.g. Of course, the results are rather speculative and based on a small sample…

Move 5- Summarizing the study

Author makes a brief summary of the main points from the perspective of the study.

e.g. In summary, the research presented in this paper offers a contrastive textlinguistics study of rhetorical differences between texts …

Move 6- Evaluating the study

This move functions to evaluate the overall study by pointing out the limitations, indicating the contributions or evaluating the methodology.

Step 1- Indicating limitations

e.g. The present study has raised a number of interesting differences, but a larger corpus is needed to establish how far they can be generalized…

Step 2- Indicating the significance

e.g. What is new in our study is the links we try to find with school performance, and the within family dynamics of the accommodation process…

Step 3- Evaluating the methodology

e.g. She performed extremely well in the experiment (as well as in the Japanese course), but it is questionable whether her experimental data represent the strategy she would employ outside of the laboratory …

Move 7- Deductions from the research

Author uses this move to extend beyond the results by suggesting what can be done to solve the problems identified by the research, pointing out the line of further study or drawing pedagogic implications.

Step 1- Making suggestions

e.g. Where such complex methods are used it may be better for the writer to provide a full and specific description of…

Step 2- Recommending further research

e.g. Further research might be profitably conducted within a single discipline to determine the degree variability according to subdiscipline, ideology, region of origin and level of prestige …

Step 3- Drawing pedagogic implications

e.g. The findings of this study may have some implications for the teaching of EAP…

C.   Eveyik-Aydin, Karabacak & Akyel's Framework Proposed for RA Discussions with Conclusions

(M1) Move 1- Setting Background

    (M1S1) Step 1: Providing methodological information regarding the study

    (M1S2) Step 2: Providing conceptual background by

        A) describing the established knowledge and stating the gaps

        B) making generalizations and claims based on the established knowledge

    (M1S3) Step 3: Making statements about the organization of the information to be presented

(M2) Move 2- Reporting Selected Results

    (M2S1) Step 1: indicating whether or not results support the hypotheses/expectations

    (M2S2) Step 2: providing relevant statistics and/or examples for a specific result

    (M2S3) Step 3: giving integrated results to summarize the main findings

(M3) Move 3- Discussing Results

    (M3S1) Step 1: Interpreting results

        A) stating what the results suggest and making claims/ arguments based on the findings of the study

        B) referring to established knowledge on topic, to other authors' views and/or to the findings of other studies in the literature

    (M3S2) Step 2: Explaining reasons for results

        A) providing reasons for the (un)expected results

        B) referring to other authors' explanations and/or research findings in the literature

        C) evaluating the explanations provided (by other studies)

(M3S3) Step 3: Comparing with results of previous work

  A) stating whether or not the results support the findings/claims/hypotheses of other authors

  B) giving information about the results of other studies / about other views

  C) by making comments on the results of other studies

 (M3S4) Step 4: Evaluating the Results

(M4) Move 4- Summarizing the study

 (M4S1) Step 1: by reminding what has been done in this study

 (M4S2) Step 2: by reminding what has been suggested by the study

(M5) Move 5- Evaluating the study

 (M5S1) Step 1- Acknowledging limitations

 (M5S2) Step 2- Indicating the significance/contribution of the study

 (M5S3) Step 3- Justifying the methodology

(M6) Move 6- Drawing Implications

 (M6S1) Step 1: Implications for Future Research

  A) Stating and justifying the need for future research (by raising questions or pointing to the issues to be considered)

  B) Making methodological suggestions for future research (by occasionally referring to other authors' views in the literature)

 (M6S2) Step 2: Pedagogical implications

  A) deducted by the author referring to main findings of the study

  B) referring to established knowledge or other authors to seek support for the implications to be drawn.

	ESP 2009_24_QUAL					Coder 2																				
Row Total		M1S1	M1S2A	M1S2B	M1S3	M2S1	M2S2	M2S3	M3S1A	M3S1B	M3S2A	M3S2B	M3S2C	M3S3A	M3S3B	M3S3C	M3S4	M4S1	M4S2	M5S1	M5S2	M5S3	M6S1A	M6S1B	M6S2A	M6S2B
4	M1S1	4																								
5	M1S2A		3			1											1									
0	M1S2B																									
1	M1S3				1																					
0	M2S1																									
12	M2S2						9		2									1								
0	M2S3																									
10	M3S1A						1		7		1							1								
0	M3S1B																									
1	M3S2A										1															
13	M3S2B		1								1	10	1													
2	M3S2C												2													
2	M3S3A													2												
1	M3S3B														1											
0	M3S3C																									
3	M3S4																3									
1	M4S1																	1								
1	M4S2						1												1							
1	M5S1																			1						
0	M5S2																									
0	M5S3																									
3	M6S1A																						3			
3	M6S1B																							3		
5	M6S2A																								5	
1	M6S2B																									1
Column Total		4	4	0	1	0	11	0	10	0	3	10	3	2	1	1	5	1	1	1	0	0	3	3	5	1
		0,2286	0,2857	0	0,0143	0	1,8857	0	1,4286	0	0,0429	1,8571	0,0857	0,0857	0,0143		0,21429	0,0143	0,0143	0,0143			0,1286	0,1286	0,3571	0,0143

$f_c = 6,7714$

$F_o = 58$

$N = 70$

$$K = \frac{f_o - f_c}{N - f_c}$$

$K = 0,8102$

*In order to calculate K, the move –step categories of 70 sentences (N=70) independently coded by two raters were entered into this table in Excel. The frequency of *agreements* between the raters were placed in the diagonal cells (e.g., both raters agreed that 9 sentences had M2S2). *Disagreements* between the coders, however, were placed in off-diagonal cells (e.g., while coder 1 interpreted two sentences as having M2S2, the second coder identified M3S1A in these sentences). In this table f_o is equal to the sum of values in the diagonal cells (f_o =58), which means that the coders showed agreement on 58 of the 70 sentences. To compute the expected frequency of agreements by chance, on the other hand, marginal values (sum of values in each row and column) were computed and placed in the formula $\frac{row\ total * column\ total}{N}$ for each diagonal cell (e.g. expected frequency for M1S2A is $\frac{5*4}{70} = 0,2857$). F_c is the sum of all expected frequencies calculated for each of the diagonal cells shown in the table (f_c=6,7714). Then using the formula $K = \frac{f_o - f_c}{N - f_c}$, K coefficient of 0,81 was found: $K = \frac{58 - 6,77}{70 - 6,77} = 0,81$

E. Move-Step Categories Identified in Each RA within the Quantitative and Qualitative Subcorpus of the Study

Quantitative Subcorpus

QUAN DATA	M1S1	M1S2A	M1S2B	M1S3	M2S1	M2S2	M2S3	M3S1A	M3S1B	M3S2A	M3S2B	M3S2C	M3S3A	M3S3B	M3S3C	M3S4	M4S1	M4S2	M5S1	M5S2	M5S3	M6S1A	M6S1B	M6S2A	M6S2B
RA1	X	X	X		X	X		X	X				X			X	X	X	X		X	X	X	X	
RA2	X	X	X		X	X		X					X		X	X		X	X			X	X	X	
RA3	X	X		X		X		X	X	X	X		X	X	X	X	X	X	X			X	X		X
RA4	X		X	X		X	X	X	X	X			X	X		X	X	X	X	X	X	X	X	X	X
RA5	X	X		X		X	X	X		X	X		X	X	X	X		X	X			X	X	X	
RA6	X	X	X		X			X	X	X	X	X	X			X			X	X	X	X	X	X	X
RA7	X	X				X	X	X	X	X	X	X	X			X		X	X			X	X	X	X
RA8	X	X			X	X		X	X	X	X		X			X		X	X			X	X		
RA9	X	X	X	X		X	X	X	X	X			X	X	X	X						X	X	X	X
RA10	X					X		X	X	X			X	X	X	X		X		X		X	X	X	X
RA11	X	X	X			X		X	X	X			X	X	X	X	X		X	X	X	X	X	X	
RA12	X	X				X	X	X	X	X			X	X		X	X	X	X			X	X	X	
RA13	X	X		X		X		X	X				X					X	X	X		X	X	X	X
RA14	X	X				X	X	X					X	X				X	X	X			X	X	X
RA15	X	X	X			X	X	X		X	X		X	X		X	X		X	X	X	X	X	X	
RA16	X		X	X		X	X	X		X	X		X			X		X	X	X	X	X	X	X	X
RA17	X	X				X	X	X		X	X		X	X		X	X	X	X	X		X	X	X	X
RA18	X	X		X		X		X	X	X	X	X	X	X	X			X	X			X	X	X	X
# of RAs	18	15	8	5	6	17	9	18	12	14	9	3	18	11	7	15	9	14	15	9	7	17	18	16	11
% of RAs	100	83,333	44,444	27,778	33,333	94,444	50	100	66,667	77,778	50	16,667	100	61,111	38,889	83,333	50	77,778	83,333	50	38,889	94,444	100	88,889	61,111

The marked (X) categories were observed at least once in these RAs.

QUAL Data

	M1S1	M1S2A	M1S2B	M1S3	M2S1	M2S2	M2S3	M3S1A	M3S1B	M3S2A	M3S2B	M3S2C	M3S3A	M3S3B	M3S3C	M3S4	M4S1	M4S2	M5S1	M5S2	M5S3	M6S1A	M6S1B	M6S2A	M6S2B
RA19	X	X		X		X		X	X					X		X	X	X	X			X		X	X
RA20	X					X		X		X			X	X		X	X	X	X		X	X	X	X	X
RA21		X	X	X		X	X	X	X		X		X			X		X		X		X	X	X	X
RA22	X	X		X		X		X		X	X		X	X	X	X	X	X		X		X			
RA23		X				X		X	X	X		X	X			X	X		X	X	X	X	X	X	X
RA24	X	X		X		X		X		X	X	X	X	X	X	X	X	X	X			X	X	X	X
RA25		X	X			X		X	X	X			X				X		X	X				X	X
RA26		X	X	X		X		X	X				X	X		X		X				X		X	X
RA27								X		X	X		X					X						X	X
RA28	X	X		X		X	X	X		X							X	X				X		X	X
RA29	X	X	X	X		X	X	X	X	X			X	X		X	X	X		X		X	X	X	X
RA30		X	X	X		X	X	X	X				X			X		X				X	X	X	X
RA31		X	X					X									X	X	X			X	X	X	X
RA32		X				X		X	X	X	X							X		X		X	X	X	X
RA33	X	X				X		X	X	X			X	X		X		X	X					X	X
RA34	X	X					X	X					X					X	X	X		X	X	X	
RA35	X																		X	X		X		X	X
RA36	X	X		X		X		X	X									X				X		X	X
Total # RA	10	15	6	9	0	14	5	17	10	10	5	2	12	7	2	10	8	15	8	7	2	15	9	17	16
%	55,56	83,33	33,333	50	0	77,78	27,78	94,444	55,5556	55,556	27,7778	11,111	66,667	38,89	11,11	55,56	44,4	83,3	44,4	38,9	11,1	83,33	50	94,44	88,889

The marked (X) categories were observed at least once in these RAs.

F. The Number of Move Occurrences at Step/Substep Level in Each RA

Number of occurrences of each move in each RA

Articles	M1	M2	M3	M4	M5	M6	M1S1	M1S2A	M1S2B	M1S3	M2S1	M2S2	M2S3	M3S1A	M3S1B	M3S2A	M3S2B	M3S2C	M3S3A	M3S3B	M3S3C	M3S4	M4S1	M4S2	M5S1	M5S2	M5S3	M6S1A	M6S1B	M6S2A	M6S2B
RA1	29	9	16	7	3	9	5	17	7	0	1	8	0	10	1	0	0	0	3	0	0	2	3	4	2	0	1	5	3	1	0
RA2	5	5	4	1	5	11	1	2	2	0	1	4	0	1	0	0	0	0	1	0	1	1	0	1	5	0	0	2	3	6	0
RA3	14	16	44	6	7	13	4	9	0	1	0	15	1	1	1	13	5	0	5	9	8	2	3	3	7	0	0	5	7	0	1
RA4	7	3	27	2	16	19	4	0	1	2	0	2	1	9	2	10	0	0	1	2	0	3	1	1	7	5	4	3	2	12	2
RA5	10	6	25	1	8	8	6	3	0	1	0	4	2	2	0	12	2	0	4	3	1	1	0	1	8	0	0	3	2	3	0
RA6	17	3	24	1	6	21	8	7	2	0	3	0	0	6	5	3	4	1	3	0	0	2	1	0	1	2	3	1	13	7	0
RA7	7	18	27	1	6	32	1	6	0	0	0	16	2	8	4	4	3	1	5	0	0	2	0	1	4	0	2	2	4	17	9
RA8	6	7	21	3	1	8	3	3	0	0	3	4	0	9	1	4	4	0	2	0	0	1	0	3	1	0	0	5	3	0	0
RA9	24	10	16	0	0	20	8	12	2	2	0	7	3	4	1	1	0	0	5	1	2	2	0	0	0	0	0	3	6	8	3
RA10	7	28	35	7	1	33	7	0	0	0	0	28	0	11	1	1	0	0	13	6	1	2	0	7	0	1	0	6	5	14	8
RA11	12	11	33	3	6	16	6	3	3	0	0	11	0	8	1	1	0	0	12	5	4	2	3	0	1	1	4	6	8	2	0
RA12	12	12	15	8	2	12	1	11	0	0	0	11	1	6	1	2	0	0	3	2	0	1	3	5	2	0	0	4	6	2	0
RA13	9	2	5	2	2	8	5	4	0	0	1	1	0	1	1	0	0	0	3	0	0	0	0	2	1	1	0	3	2	2	1
RA14	3	7	10	5	4	19	1	2	0	0	0	3	4	5	0	0	0	0	4	1	0	0	0	5	2	2	0	0	3	7	9
RA15	12	6	32	1	17	12	6	5	1	0	0	4	2	11	0	3	12	0	2	2	0	2	1	0	12	3	2	5	1	6	0
RA16	6	13	14	2	9	18	2	0	1	3	0	12	1	4	0	6	1	0	2	0	0	1	1	1	2	2	5	2	2	6	8
RA17	7	4	32	4	5	33	5	2	0	0	0	3	1	19	0	2	3	0	2	5	0	1	2	2	3	2	0	6	3	20	4
RA18	7	4	22	3	0	15	5	2	0	0	3	1	0	4	3	2	4	1	5	1	2	0	2	1	0	0	0	7	1	5	2
QUAN Ms	194	164	402	57	98	307	78	88	19	9	12	134	18	119	22	64	38	3	75	37	19	25	20	37	58	19	21	68	74	118	47
RA19	6	9	21	7	3	15	1	3	0	2	0	9	0	8	10	0	0	0	1	2	0	0	1	6	3	0	0	2	2	6	5
RA20	2	1	15	4	8	15	2	0	0	0	0	1	0	8	0	1	0	0	3	2	0	1	2	2	7	0	1	1	1	9	4
RA21	9	10	18	6	2	21	0	4	4	1	0	7	3	6	7	0	1	0	2	0	0	2	0	6	0	2	0	3	2	12	4
RA22	9	17	25	6	1	2	7	1	0	1	0	17	0	12	0	5	1	0	3	3	1	0	3	3	0	1	0	2	0	0	0
RA23	9	11	22	1	12	13	0	9	0	0	0	11	0	7	1	8	0	2	4	0	0	0	1	0	3	6	3	2	1	9	1
RA24	8	13	33	4	1	12	3	4	0	1	0	13	0	8	0	3	10	3	2	1	1	5	1	3	1	0	0	3	3	5	1
RA25	6	12	30	3	2	16	0	4	2	0	0	12	0	19	5	4	0	0	1	0	0	1	0	3	1	1	0	0	0	11	5
RA26	5	7	26	0	0	22	0	3	1	1	0	7	0	14	2	1	0	0	4	4	0	1	0	0	0	0	0	1	0	18	3
RA27	0	0	20	3	0	4	0	0	0	0	0	0	0	16	0	2	1	0	1	0	0	1	0	3	0	0	0	0	0	2	2
RA28	9	37	11	4	0	15	5	3	0	1	0	36	1	9	0	1	0	0	0	0	0	1	0	4	0	0	0	1	0	9	5
RA29	27	6	25	8	1	22	3	15	8	1	0	5	1	15	1	2	0	0	5	1	0	1	1	7	0	1	0	7	1	12	2
RA30	51	5	22	4	0	10	0	31	19	1	0	4	1	12	4	1	0	0	3	0	0	2	0	4	0	0	0	3	6	1	0
RA31	11	0	2	2	0	22	0	8	2	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	9	1	11	1
RA32	2	1	6	2	5	17	0	2	0	0	0	1	0	1	1	2	2	0	0	0	0	0	1	1	5	0	0	1	1	12	3
RA33	13	16	17	8	0	14	10	3	0	0	0	16	0	10	2	1	0	0	1	1	0	2	0	8	0	0	0	0	0	13	1
RA34	5	1	9	3	6	13	2	3	0	0	0	0	1	5	0	0	0	0	4	0	0	0	0	3	3	3	0	2	10	1	0
RA35	3	0	0	0	4	41	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	12	0	23	6
RA36	6	2	30	1	0	15	1	3	0	2	0	2	0	21	9	0	0	0	0	0	0	0	0	1	0	0	0	2	0	11	2
QUAL Ms	181	148	332	66	45	289	37	96	36	12	0	141	7	172	42	31	15	5	34	14	2	17	11	55	24	17	4	51	28	165	45
Total #	375	312	734	123	143	596	115	184	55	21	12	275	25	291	64	95	53	8	109	51	21	42	31	92	82	36	25	119	102	283	92

G.   The Percentage of Move Occurrences at Step/Substep Level in Each RA

Article	M1	M2	M3	M4	M5	M6	M1S1	M1S2A	M1S2B	M1S3	M2S1	M2S2	M2S3	M3S1A	M3S1B	M3S2A	M3S2B	M3S2C	M3S3A	M3S3B	M3S3C	M3S4	M4S1	M4S2	M5S1	M5S2	M5S3	M6S1A	M6S1B	M6S2A	M6S2I	%
RA1	39,73	12,33	21,92	9,589	4,11	12,33	6,849	23,288	9,589	0	1,37	10,96	0	13,6986	1,3699	0	0	0	4,10959	0	0	2,74	4,11	5,479	2,74	0	1,37	6,84932	4,1096	1,3699	0	100
RA2	16,13	16,13	12,9	3,226	16,13	35,48	3,226	6,4516	6,4516	0	3,226	12,9	0	3,22581	0	0	0	0	3,22581	0	3,22581	3,226	0	3,226	16,13	0	0	6,45161	9,6774	19,355	0	100
RA3	14	16	44	6	7	13	4	9	0	1	0	15	1	1	1	13	5	0	5	9	8	2	3	3	7	0	0	5	7	0	1	100
RA4	9,459	4,054	36,49	2,703	21,62	25,68	5,405	0	1,3514	2,703	0	2,703	1,351	12,1622	2,7027	13,514	0	0	1,35135	2,7027	0	4,054	1,351	1,351	9,459	6,757	5,405	4,05405	2,7027	16,216	2,703	100
RA5	17,24	10,34	43,1	1,724	13,79	13,79	10,34	5,1724	0	1,724	0	6,897	3,448	3,44828	0	20,69	3,4483	0	6,89655	5,17241	1,72414	1,724	0	1,724	13,79	0	0	5,17241	3,4483	5,1724	0	100
RA6	23,61	4,167	33,33	1,389	8,333	29,17	11,11	9,7222	2,7778	0	4,167	0	0	8,33333	6,9444	4,1667	5,5556	1,38889	4,16667	0	0	2,778	1,389	0	1,389	2,778	4,167	1,38889	18,056	9,7222	0	100
RA7	7,692	19,78	29,67	1,099	6,593	35,16	1,099	6,5934	0	0	0	17,58	2,198	8,79121	4,3956	4,3956	3,2967	1,0989	5,49451	0	0	2,198	0	1,099	4,396	0	2,198	2,1978	4,3956	18,681	9,89	100
RA8	13,04	15,22	45,65	6,522	2,174	17,39	6,522	6,5217	0	0	6,522	8,696	0	19,5652	2,1739	8,6957	8,6957	0	4,34783	0	0	2,174	0	6,522	2,174	0	0	10,8696	6,5217	0	0	100
RA9	34,29	14,29	22,86	0	0	28,57	11,43	17,143	2,8571	2,857	0	10	4,286	5,71429	1,4286	1,4286	0	0	7,14286	1,42857	2,85714	2,857	0	0	0	0	0	4,28571	8,5714	11,429	4,286	100
RA10	6,306	25,23	31,53	6,306	0,901	29,73	6,306	0	0	0	0	25,23	0	9,90991	0,9009	0,9009	0	0	11,7117	5,40541	0,9009	1,802	0	6,306	0	0,901	0	5,40541	4,5045	12,613	7,207	100
RA11	14,81	13,58	40,74	3,704	7,407	19,75	7,407	3,7037	3,7037	0	0	13,58	0	9,87654	1,2346	1,2346	0	0	14,8148	6,17284	4,93827	2,469	3,704	0	1,235	1,235	4,938	7,40741	9,8765	2,4691	0	100
RA12	19,67	19,67	24,59	13,11	3,279	19,67	1,639	18,033	0	0	0	18,03	1,639	9,83607	1,6393	3,2787	0	0	4,91803	3,27869	0	1,639	4,918	8,197	3,279	0	0	6,55738	9,8361	3,2787	0	100
RA13	32,14	7,143	17,86	7,143	7,143	28,57	17,86	14,286	0	0	3,571	3,571	0	3,57143	3,5714	0	0	0	10,7143	0	0	0	0	7,143	3,571	3,571	0	10,7143	7,1429	7,1429	3,571	100
RA14	6,25	14,58	20,83	10,42	8,333	39,58	2,083	4,1667	0	0	0	6,25	8,333	10,4167	0	0	0	0	8,33333	2,08333	0	0	0	10,42	4,167	4,167	0	0	6,25	14,583	18,75	100
RA15	15	7,5	40	1,25	21,25	15	7,5	6,25	1,25	0	0	5	2,5	13,75	0	3,75	15	0	2,5	2,5	0	2,5	1,25	0	15	3,75	2,5	6,25	1,25	7,5	0	100
RA16	9,677	20,97	22,58	3,226	14,52	29,03	3,226	0	1,6129	4,839	0	19,35	1,613	6,45161	0	9,6774	1,6129	0	3,22581	0	0	1,613	1,613	1,613	3,226	3,226	8,065	3,22581	3,2258	9,6774	12,9	100
RA17	8,235	4,706	37,65	4,706	5,882	38,82	5,882	2,3529	0	0	0	3,529	1,176	22,3529	0	2,3529	3,5294	0	2,35294	5,88235	0	1,176	2,353	2,353	3,529	2,353	0	7,05882	3,5294	23,529	4,706	100
RA18	13,73	7,843	43,14	5,882	0	29,41	9,804	3,9216	0	0	5,882	1,961	0	7,84314	5,8824	3,9216	7,8431	1,96078	9,80392	1,96078	3,92157	0	3,922	1,961	0	0	0	13,7255	1,9608	9,8039	3,922	100
RA19	9,836	14,75	34,43	11,48	4,918	24,59	1,639	4,918	0	3,279	0	14,75	0	13,1148	16,393	0	0	0	1,63934	3,27869	0	0	1,639	9,836	4,918	0	0	3,27869	3,2787	9,8361	8,197	100
RA20	4,444	2,222	33,33	8,889	17,78	33,33	4,444	0	0	0	0	2,222	0	17,7778	0	2,2222	0	0	6,66667	4,44444	0	2,222	4,444	4,444	15,56	0	2,222	2,22222	2,2222	20	8,889	100
RA21	13,64	15,15	27,27	9,091	3,03	31,82	0	6,0606	6,0606	1,515	0	10,61	4,545	9,09091	10,606	0	1,5152	0	3,0303	0	0	3,03	0	9,091	0	3,03	0	4,54545	3,0303	18,182	6,061	100
RA22	15	28,33	41,67	10	1,667	3,333	11,67	1,6667	0	1,667	0	28,33	0	20	0	8,3333	1,6667	0	5	5	1,66667	0	5	5	0	1,667	0	3,33333	0	0	0	100
RA23	13,24	16,18	32,35	1,471	17,65	19,12	0	13,235	0	0	0	16,18	0	10,2941	1,4706	11,765	0	2,94118	5,88235	0	0	0	1,471	0	4,412	8,824	4,412	2,94118	1,4706	13,235	1,471	100
RA24	11,27	18,31	46,48	5,634	1,408	16,9	4,225	5,6338	0	1,408	0	18,31	0	11,2676	0	4,2254	14,085	4,22535	2,8169	1,40845	1,40845	7,042	1,408	4,225	1,408	0	0	4,22535	4,2254	7,0423	1,408	100
RA25	8,696	17,39	43,48	4,348	2,899	23,19	0	5,7971	2,8986	0	0	17,39	0	27,5362	7,2464	5,7971	0	0	1,44928	0	0	1,449	0	4,348	1,449	1,449	0	0	0	15,942	7,246	100
RA26	8,333	11,67	43,33	0	0	36,67	0	5	1,6667	1,667	0	11,67	0	23,3333	3,3333	1,6667	0	0	6,66667	6,66667	0	1,667	0	0	0	0	0	1,66667	0	30	5	100
RA27	0	0	74,07	11,11	0	14,81	0	0	0	0	0	0	0	59,2593	0	7,4074	3,7037	0	3,7037	0	0	0	0	11,11	0	0	0	0	0	7,4074	7,407	100
RA28	11,84	48,68	14,47	5,263	0	19,74	6,579	3,9474	0	1,316	0	47,37	1,316	11,8421	0	1,3158	0	0	0	0	0	1,316	0	5,263	0	0	0	1,31579	0	11,842	6,579	100
RA29	30,34	6,742	28,09	8,989	1,124	24,72	3,371	16,854	8,9888	1,124	0	5,618	1,124	16,8539	1,1236	2,2472	0	0	5,61798	1,1236	0	1,124	1,124	7,865	0	1,124	0	7,86517	1,1236	13,483	2,247	100
RA30	55,43	5,435	23,91	4,348	0	10,87	0	33,696	20,652	1,087	0	4,348	1,087	13,0435	4,3478	1,087	0	0	3,26087	0	0	2,174	0	4,348	0	0	0	3,26087	6,5217	1,087	0	100
RA31	29,73	0	5,405	5,405	0	59,46	0	21,622	5,4054	2,703	0	0	0	2,7027	0	0	0	0	0	0	0	2,703	2,703	2,703	0	0	0	24,3243	2,7027	29,73	2,703	100
RA32	6,061	3,03	18,18	6,061	15,15	51,52	0	6,0606	0	0	0	3,03	0	3,0303	3,0303	6,0606	6,0606	0	0	0	0	0	3,03	3,03	15,15	0	0	3,0303	3,0303	36,364	9,091	100
RA33	19,12	23,53	25	11,76	0	20,59	14,71	4,4118	0	0	0	23,53	0	14,7059	2,9412	1,4706	0	0	1,47059	1,47059	0	2,941	0	11,76	0	0	0	0	0	19,118	1,471	100
RA34	13,51	2,703	24,32	8,108	16,22	35,14	5,405	8,1081	0	0	0	0	2,703	13,5135	0	0	0	0	10,8108	0	0	0	0	8,108	8,108	8,108	0	5,40541	27,027	2,7027	0	100
RA35	6,25	0	0	0	8,333	85,42	6,25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2,083	6,25	0	25	47,917	12,5		100
RA36	11,11	3,704	55,56	1,852	0	27,78	1,852	5,5556	0	3,704	0	3,704	0	38,8889	16,667	0	0	0	0	0	0	0	0	1,852	0	0	0	3,7037	0	20,37	3,704	100

H.   Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
M1	QUAN	18	16,72	9,831	2,317
	QUAL	18	14,88	12,728	3,000
M2	QUAN	18	12,97	6,217	1,465
	QUAL	18	12,10	12,563	2,961
M3	QUAN	18	31,60	10,270	2,421
	QUAL	18	31,74	17,716	4,176
M4	QUAN	18	4,89	3,579	,844
	QUAL	18	6,32	3,842	,906
M5	QUAN	18	8,25	6,689	1,577
	QUAL	18	5,01	6,798	1,602
M6	QUAN	18	25,56	8,933	2,106
	QUAL	18	29,94	19,383	4,569
M1S1	QUAN	18	6,76	4,219	,994
	QUAL	18	3,34	4,326	1,020
M1S2A	QUAN	18	7,59	6,641	1,565
	QUAL	18	7,92	8,593	2,025
M1S2B	QUAN	18	1,64	2,662	,628
	QUAL	18	2,54	5,250	1,237
M1S3	QUAN	18	,73	1,399	,330
	QUAL	18	1,08	1,206	,284
M2S1	QUAN	18	1,37	2,243	,529
	QUAL	18	,00	,000	,000
M2S2	QUAN	18	10,07	7,029	1,657
	QUAL	18	11,50	12,514	2,950
M2S3	QUAN	18	1,53	2,144	,505
	QUAL	18	,60	1,229	,290
M3S1A	QUAN	18	9,44	5,542	1,306
	QUAL	18	17,01	13,995	3,299
M3S1B	QUAN	18	1,85	2,107	,497
	QUAL	18	3,73	5,489	1,294
M3S2A	QUAN	18	5,06	5,819	1,372
	QUAL	18	2,98	3,508	,827
M3S2B	QUAN	18	3,00	4,158	,980
	QUAL	18	1,50	3,542	,835
M3S2C	QUAN	18	,25	,588	,139
	QUAL	18	,40	1,180	,278
M3S3A	QUAN	18	6,12	3,670	,865

		18			
	QUAL	18	3,22	3,052	,719
M3S3B	QUAN	18	2,53	2,752	,649
	QUAL	18	1,30	2,100	,495
M3S3C	QUAN	18	1,42	2,298	,542
	QUAL	18	,17	,499	,118
M3S4	QUAN	18	1,94	1,116	,263
	QUAL	18	1,43	1,810	,427
M4S1	QUAN	18	1,53	1,720	,405
	QUAL	18	1,16	1,632	,385
M4S2	QUAN	18	3,36	3,198	,754
	QUAL	18	5,17	3,730	,879
M5S1	QUAN	18	5,06	5,155	1,215
	QUAL	18	2,95	5,046	1,189
M5S2	QUAN	18	1,60	2,028	,478
	QUAL	18	1,69	2,935	,692
M5S3	QUAN	18	1,59	2,466	,581
	QUAL	18	,37	1,136	,268
M6S1A	QUAN	18	5,92	3,414	,805
	QUAL	18	5,34	7,302	1,721
M6S1B	QUAN	18	6,23	4,025	,949
	QUAL	18	3,04	6,277	1,480
M6S2A	QUAN	18	9,59	6,969	1,643
	QUAL	18	16,90	12,675	2,988
M6S2B	QUAN	18	3,83	5,318	1,253
	QUAL	18	4,67	3,743	,882

I.   Cycles at Move Level

2-cycle Patterns at Move Level

Moves	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f of occurrence	*N of RAs*	*Normalized f**	*f of occurrence*	*N of RAs*	*Normalized f**
M2 M3	72	18	**59**	52	12	**49**
M3 M1	41	17	**34**	32	14	**30**
M3 M2	39	13	**32**	39	12	**37**
M3 M6	37	17	**30**	37	15	**35**
M1 M3	34	15	**28**	25	11	**24**
M1 M2	27	14	**22**	16	8	**15**
M5 M6	26	13	**21**	12	6	**11**
M6 M5	21	12	**17**	11	6	**10**
M6 M3	18	12	**15**	20	12	**19**
M4 M6	17	11	**14**	11	10	**10**
M6 M4	12	8	**10**	14	10	**13**

*Numbers rounded to the nearest digit

3- cycle Patterns at Move Level

Moves	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f of occurrence	*N of RAs*	*Normalized f**	*f of occurrence*	*N of RAs*	*Normalized f**
M3-M2-M3	30	12	**25**	36	12	**34**
M2-M3-M2	-	-	**-**	16	8	**15**
M1-M2-M3	21	12	**17**	10	5	**9**
M6-M3-M6	12	8	**10**	13	6	**12**
M3-M1-M3	18	12	**15**	11	7	**10**
M1-M3-M2	16	9	**13**	5	4	**5**
M3-M1-M2	14	8	**12**	6	3	**6**
M6-M5-M6	14	10	**12**	7	5	**7**

4- cycle Patterns at Move Level

Moves	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f of occur.	*N of RAs*	*Normalized f **	*f of occur.*	*N of RAs*	*Normalized f**
M3-M2-M3-M2	9	7	**7**	15	8	**14**
M3-M1-M2-M3	14	8	**12**	-	-	**-**

5- cycle Patterns at Move Level

Move Cycles	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f of occur.	*N of RAs*	*Normalized f **	*f of occur.*	*N of RAs*	*Normalized f**
M3-M2-M3-M2-M3	6	5	**5**	13	7	**12**

J.   Cycles at Move-Step Level

2- cycle Patterns at Step Level

Steps	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f	*N of RAs*	*Normalized f**	*f*	*N of RAs*	*Normalized f**
M6S2-M6S2	*38*	*9*	**31**	*38*	*14*	**36**
M2S2-M3S1	25	12	**20**	35	11	**33**
M6S1-M6S1	25	13	**20**	11	7	**10**
M2S2-M3S3	22	10	**18**	-	-	**-**
M2S1-M3S1	12	6	**10**	-	-	**-**
M3S1-M3S2	20	9	**16**	5	3	**5**
M5S1-M6S1	15	11	**12**	9	5	**8**
M3S1-M2S2	14	8	**11**	23	10	**22**
M3S1-M1S2	11	8	**9**	13	9	**12**
M3S1-M6S2	10	7	**8**	20	12	**19**
M1S2-M1S2	9	4	**7**	26	6	**25**
M3S1-M3S1	9	6	**7**	22	8	**21**
M6S2-M3S1	7	4	**6**	13	7	**12**
M1S2-M3S1	5	4	**4**	13	7	**12**
M4S2-M6S2	15	9	**12**	7	7	**7**

*Numbers rounded to the nearest digit

3-cycle Patterns at Step Level

Steps	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f	*N of RAs*	*Normalized f **	*f*	*N of RAs*	*Normalized f**
M6S2-M6S2-M6S2	22	7	**18**	19	10	**18**
M1S2-M1S2-M1S2	-	-	**-**	15	3	**14**
M2S2-M3S1-M2S2	7	6	**6**	13	7	**12**
M2S2-M3S1-M3S2	5	4	**4**	13	7	**12**
M3S1-M2S2-M3S1	-	-	**-**	17	7	**16**

4- cycle Patterns at Step Level

Steps	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f	*N of RAs*	*Normalized f**	*f*	*N of RAs*	*Normalized f**
M6S2-M6S2-M6S2-M6S2	15	5	**12**	7	4	**7**

*Numbers rounded to the nearest digit

K.   Cycles at Substep Level

2- cycle Patterns at Substep Level

Substeps	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f	*N of RAs*	*Normalized f **	*f*	*N of RAs*	*Normalized f**
M2S2-M3S1A	22	11	18	31	11	29
M6S2B- M6S2A	20	8	16	18	11	17
M2S2-M3S3A	18	9	15	-	-	-
M6S2A-M6S2B	18	8	15	20	13	19
M3S3A-M3S3B	15	11	12	8	5	8
M3S1A-M2S2	13	7	11	23	10	22
M1S1-M2S2	15	7	12	7	3	7
M5S1-M6S1B	15	10	12	6	4	6
M1S2A-M1S2B	7	3	6	16	6	15
M3S1A-M6S2A	6	4	5	13	8	12

3- cycle Patterns at Substep Level

Substeps	QUAN RAs (N=18)			QUAL RAs (N=18)		
	f	*N of RAs*	*Normalized f **	*f*	*N of RAs*	*Normalized f**
M6S2A-M6S2B-M6S2A	12	6	**10**	13	8	**12**
M3S1A-M2S2-M3S1A	-	-	**-**	15	8	**14**

*Numbers rounded to the nearest digit

L.   The Size of 6 Moves Identified in the Study

Corpus	Word Types (Different words)	Word Tokens (Words in text)
Whole Corpus	5,324	66272
QUAN Corpus	3,588	34767
QUAL Corpus	3,670	31505
M1	2,275	10225
M1 QUAN	1,445	5,050
M1 QUAL	1,442	5,175
M2	1,708	8,150
M2 QUAN	993	4,249
M2 QUAL	1,109	3,901
M3	3,297	22,626
M3 QUAN	2,097	12,430
M3 QUAL	2,157	10,196
M4	1,029	3,382
M4 QUAN	574	1,491
M4 QUAL	658	1,891
M5	1,077	3,829
M5 QUAN	794	2,479
M5 QUAL	527	1,350
M6	2,844	18,060
M6 QUAN	1,843	9,068
M6 QUAL	1,885	8,992

M.   Formulaic Bundles Identified in Move 1

5-PoS tag Bundle in M1	**f**	**Expected**	**Log Likelihood**
nn vbd dt nn in	**9**	**4**	**5.66475**
(noun, verb past tense, determiner, noun, preposition or subordinating conjunction)			

M1S1
second *question examined the amount of* student   (RA10_8 M1S1)
research *question examined the relationship between* student (RA10_45 M1S1)
present *research examined the usefulness of* two   (RA3_3 M1S1)
the *researcher addressed the role of* motivation   (RA10_41 M1S1)
this *study concerned the relationship between* feedback   (RA3_28 M1S1)
this *study investigated the relationship between* output   (RA12_3 M1S1)
This *study examined the relevance of* the   (RA14_1 M1S1)
theoretical *underpining was that progress in* listening   (RA4_35 M1S1)
M1S2
None discussed the procedure of how   (RA23_48 M1S2A)

4-PoS tag Bundle in M1	**f**	**Expected**	**Log Likelihood**
nn vbd dt nn	**11**	**7**	**5.85946**
(noun-verb past tense- determiner-noun)			

M1S1
second *question concerned the extent* to w   (RA15_31 M1S1)
*research *question examined the amount* of st   (RA10_8 M1S1)
*research *question examined the relationship* be   (RA10_45 M1S1)
*present *research examined the usefulness* of   (RA3_3 M1S1)
*the *researcher addressed the role* of m   (RA10_41 M1S1)
*this *study concerned the relationship* between   (RA3_28 M1S1)
This *study investigated the effects* of   (RA6_1 M1S1)
*this *study investigated the relationship* between   (RA12_3 M1S1)
*This *study examined the relevance* of the   (RA14_1 M1S1)
*theoretical *underpining was that progress* in   (RA4_35 M1S1)
M1S2
***None discussed the procedure* of ho   (RA23_48 M1S2A)

*Part of them was captured in other PoS bundles.

4-PoS tag Bundle in M1	f	Expected	Log Likelihood
rb vbn in jj	7	2	5.66475

(adverb- verb past participle- preposition or subordinating conjunction-adjective)

M1S1
is *often given in intensive* ESL prog   (RA9_39 M1S1)

M1S2
have *also shown that implicit* metalinguist   (RA30_C80 M1S2A)
has *predominantly focused on oral* product   (RA31_30 M1S2A)
is more *truly manifested in linguistic* perfor   (RA30_45 M1S2A)
is *usually based on linguistic* measures   (RA34_15 M1S2A)
it is *well documented that second* learners   (RA1_43 M1S2A)
has been *widely argued that formulaic* sequences   (RA1_C62 M1S2A)

5-PoS tag Bundle in M2	f	Expected	Log Likelihood
in dt jj nn,	**21**	**11**	**7.68600**
(preposition or subordinating conjunction- determiner-adjective-noun-,)			

M2S2

knew ***about the subject matter,*** the mo	(RA25_26 M2S2)
At the same time, though , ove	(RA21_35 M2S2)
, ***at the same time,*** expressed the	(RA10_7 M2S2)
herself ***as a shy person,*** did not	(RA25_30 M2S2)
that ***by the second recall,*** it had	(RA2_6 M2S2)
than ***for the traditional treatment,*** and	(RA7_18 M2S2)
score of 57.5 in the first year, to 33.5	(RA16_12 M2S2)
In this particular study, implicit	(RA30_C79 M2S2)
In the present study, the rate	(RA3_21 M2S2)
that ***in the present study,*** aspects	(RA8_C32 M2S2)
waned ***in the present study***, as it	(RA2_7 M2S2)
In the present study, 70 % of the	(RA24_10 M2S2)
In the present study, listening	(RA11_14 M2S2)
In the present study, for instan	(RA11_61 M2S2)
In the third year, the quality	(RA16_6 M2S2)
in the second year, although his	(RA16_14 M2S2)
of a statistical relationship, and	(RA28_7 M2S2)
of a statistical relationship, for	(RA28_8 M2S2)
on the other hand, the superviso	(RA19_34 M2S2)
On the other hand, the ratio	(RA24_38 M2S2)
within the Chinese program, the	(RA22_12 M2S2)

5-PoS tag Bundle in M2	f	Expected	Log Likelihood
nn vbd in dt nn	**5**	**1**	**6.63293**
(noun-verb past tense-preposition or subordinating conjunction- determiner-noun)			

M2S2

Analysis demonstrated that the heritage Spanish	(RA5_2 M2S2)
This ***analysis revealed that the number*** of	(RA16_21M2S2)
my ***informant commented on the phrase*** on	(RA28_59 M2S2)
analysis ***result revealed that all task*** type	(RA11_57 M2S2)
t-***test revealed that the study*** a	(RA10_23 M2S2)

5-PoS tag Bundle in M2	f	Expected	Log Likelihood
vbd vbn to vb jj	**5**	**0**	**6.47414**

(verb past tense- verb past participle-to-verb base form-adjective)

M2S2

students *were felt to have different* understandings	(RA26_19 M2S2)
model *was found to be useful* as	(RA24_C41 M2S2)
attempts *were made to include relevant* unpublished	(RA28_38 M2S2)
that *was theorized to be integral* to	(RA29_76 M2S2)
M2S3	
moderators *were found to be significant* predictors	(RA12_4 M2S3)

5-PoS tag Bundle in M2	f	Expected	Log Likelihood
vbd dt jj nn in	**8**	**4**	**5.8272**

(verb past tense-determiner-adjective-noun- preposition or subordinating conjunction)

M2S2

also *confirmed the textual analysis in* that	(RA23_4 M2S2)
already *had a sufficient level of* linguistic	(RA22_17 M2S2)
analysis *identified a significant relationship between* the	(RA10_46 M2S2)
who *invented a high number of* solution	(RA8_C27 M2S2)
item *revealed a significant variation in* the	(RA10_16 M2S2)
provided *suggested a varied range of* activities	(RA21_31 M2S2)
M2S3	
program *had a positive impact on* listening	(RA4_2 M2S3)
CF *had a positive effect on* the	(RA15_2 M2S3)

5-PoS tag Bundle in M2	f	Expected	Log Likelihood
nns vbd dt nn in	**6**	**2**	**5.63577**

(noun plural- verb past tense- determiner-noun- preposition or subordinating conjunction)

M2S2

all *participants expressed an interest in* the	(RA10_4 M2S2)
For instance, *students mentioned the importance of* having	(RA10_6 M2S2)
The *students reported an interest in* speaking	(RA10_2 M2S2)
A few *students said that speaking with* their	(RA10_29 M2S2)
In terms of reading, *teachers cited a range of* sources	(RA21_29 M2S2)
number of *writers boosted the statement in* some	(RA28_13 M2S2)

4-PoS tag Bundle in M2   **f**   **Expected**   **Log Likelihood**

cd nn in dt   **10**   **2**   **17.43289**

(cardinal number-noun- preposition or subordinating conjunction- determiner)

M2S2

about *one third of the* time   (RA3_29 M2S2)
over *40 % of those* that   (RA3_38 M2S2)
and *36 % of those* that   (RA3_38 M2S2)
spending *8 hr on a* single   (RA7_C78 M2S2)
for 10-*15 % of the* variance   (RA8_C32 M2S2)
approximately 10-*15 % of the* variance   (RA8_4 M3S1A + *M2S2*)
knowledge *5.5 % for the* TMS–R   (RA11_21 M2S2)
present study, *70 % of the* RAIs   (RA24_10 M2S2)
only *10 % of those* in   (RA24_10 M2S2)
as *85% of the* students   (RA33_2 M4S2 + *M2S2*)

4-PoS tag Bundle in M2   **f**   **Expected**   **Log Likelihood**

vbd dt nn in   **25**   **12**   **14.73619**

(verb past tense, determiner, noun, preposition or subordinating conjunction)

M2S2

*writers *boosted the statement in* some way   (RA28_13 M2S2)
*teachers *cited a range of* sources they   (RA21_29 M2S2)
that *expanded the content of* the uttera   (RA3_37 M2S2)
*participants *expressed an interest in* the pragm   (RA10_4 M2S2)
both *emphasized the necessity for* student   (RA19_34 M2S2)
students who *had a gain of* reported   (RA10_50 M2S2)
corpus *had an average of* 5. 2   (RA24_39 M2S2)
posttest *had an average of* 43.75 hours   (RA10_50 M2S2)
treatment *had an effect over* and above   (RA15_14 M3S1A + M2S2)
type of text also *had an impact on* foreign language   (RA12_C41 M2S2)
that *identified the error by* repeating   (RA3_38 M2S2)
analysis *illustrated a predominance of* prescri   (RA23_2 M2S2)
*students *mentioned the importance of* having   (RA10_6 M2S2)
that *provided the correction in* context   (RA3_43 M2S2)
who *read a combination of* expository a   (RA12_20 M2S2)
that *repeated the error in* combination   (RA3_14 M2S2)
students *reported an interest in* speaking   (RA10_1 M2S2)
that *reformulated the error with* no add   (RA3_37 M2S2)
that *reformulated the error within* its   (RA3_12 M2S2)
students *reported an average of* 52.36 hours   (RA10_50 M2S2)
*students *said that speaking with* their fa   (RA10_29 M2S2)
corpus *was the number of* raw tokens   (RA9_4 M2S2)
there *was a drop in* the percentage   (RA23_3 M2S2)
statement *was the estimate of* life expectanc   (RA28_15 M2S2)
there *was a relationship between* the act   (RA33_34 M2S2)

4-PoS tag Bundle in M2   **f**   **Expected**   **Log Likelihood**

4-PoS tag Bundle in M2	f	Expected	Log Likelihood
rb vbd in dt	**7**	**2**	**11.22341**

(adverb- verb, past tense- preposition or subordinating conjunction -determiner)

M2S2

They *also reflected on the* significance	(RA36_21 M2S2)
They *further claimed that the* DDL trea	(RA7_36 M2S2)
He *further explained that this* was a	(RA28_31 M2S2)
results *further showed that both* reformulatio	(RA3_11 M2S2)
above *further suggested that the* scenario	(RA19_39 M1S2A + M2S2)
*difference *quickly waned in the* present study	(RA2_7 M2S2)
requirements *simply meant that a* student had	(RA22_15 M2S2)

4-PoS tag Bundle in M2	f	Expected	Log Likelihood
vbd in dt nn	**25**	**6**	**10.73581**

(verb, past tense- preposition or subordinating conjunction-determiner-noun)

M2S2

I *asked for an example* said	(RA21_32 M2S2)
*further *claimed that the DDL* treatment wo	(RA7_36 M2S2)
*informant *commented on the phrase* on a	(RA28_59 M2S2)
He *commented that a statement* about t	(RA28_8 M2S2)
statistics *confirmed that the study* – abroad	(RA10_1 M2S2)
*Analysis *demonstrated that the heritage* Spanis	(RA5_2 M2S2)
I *found that the students* constantly	(RA25_21 M2S2)
LCP *indicated that the students* particip	(RA10_9 M2S2)
questionnaire *indicated that the students* were	(RA10_70 M2S2)
*simply *meant that a student* had the	(RA22_15 M2S2)
requirements *meant that the students* already h	(RA22_17 M2S2)
contrast , *occurred across a range* of highly_	(RA9_8 M2S2)
students *participated in these activities* for	(RA10_9 M2S2)
*also *reflected on the significance* of '	(RA36_21 M2S2)
*result *revealed that all task* types ,	(RA11_57 M2S2)
*analysis *revealed that the number* of revisi	(RA16_21 M2S2)
*t-test *revealed that the study* abroad	(RA10_23 M2S2)
students *said that the SAC* helped to	(RA33_2 M4S2 +M2S2)
*further *showed that both reformulations* and	(RA3_11 M2S2)
findings *showed that both reformulations* and	(RA3_39 M2S2)
however, *suggested that this use* by student	(RA28_8 M2S2)
*further *suggested that the scenario* tends	(RA19_39 M1S2A + M2S2)
and *waited for the learner* to supply	(RA3_38 M2S2)
Test 2 *was for the DDL* items , the	(RA7_2 M2S2)
marks *varied from a text quality* score	(RA16_12 M2S2)

4-PoS tag Bundle in M2	f	Expected	Log Likelihood
nn vbd in dt	**10**	**3**	**10.61740**

(noun- verb, past tense- preposition or subordinating conjunction-determiner)

M2S2

*Analysis demonstrated that the herita	(RA5_2 M2S2)
*Factor analysis **result revealed that all**	(RA11_57 M2S2)
*This **analysis revealed that the** number	(RA16_21 M2S2)
*my **informant commented on the** phrase	(RA28_59 M2S2)
My **informant said that this** would	(RA28_6 M2S2 + M3S2A)
the pretest **questionnaire indicated that** the	(RA10_70 M2S2)
the **staff agreed with this** as well	(RA33_3 M2S2)
of **staff agreed with this** 78 %	(RA33_3 M2S2)
*sample **t-test revealed that the** study	(RA10_23 M2S2)

M2S3

my **informant felt that the** different	(RA28_9 M2S3)

4-PoS tag Bundle in M2	f	Expected	Log Likelihood
nn vbd jj nn	**9**	**2**	**10.46617**

(noun- verb, past tense- adjective-noun)

M2S2

NNES **apprentices saw linguistic competence** playing	(RA22_24 M2S2)
study – abroad **participants made significant improvement** on	(RA10_23 M2S2)
my **participants identified sociocultural differences** as	(RA25_12 M2S2)
Reviewers took different stances be	(RA23_32 M2S2)
NNES **students brought positive influences** to	(RA22_40 M2S2)
four **students suffered marked disadvantages** because	(RA22_41 M2S2)
FL **teachers demonstrated several characteristics** of	(RA29_24 M2S2)
CF **types had differential effects** :	(RA15_16 M2S2)

M2S3

L2 **learners showed significant improvement** on	(RA5_C53 M2S3)

4-PoS tag Bundle in M2	f	Expected	Log Likelihood
vbd vbn to vb	**10**	**4**	**9.16128**

(verb, past tense-verb, past participle-to-verb, based form)

M2S2

interpretation **was found to be** a	(RA11_61M2S2)
model **was found to be** useful	(RA11_61 M2S2)
he **was relieved to observe** that	(RA28_38 M2S2)
competence **was required to become** a	(RA22_16 M2S2)
*that **was theorized to be** integral	(RA29_76 M2S2)
creativity **were found to account** only	(RA8_C32 M2S2)
moderators **were found to be** significant	(RA24_C41 M2S2)
that **were found to be** the	(RA24_C41 M2S2)
test **were found to engage** in more	(RA8_C27 M2S2)
*students **were felt to have** different	(RA26_19 M2S2)

4-PoS tag Bundle in M2   **f**   **Expected**   **Log Likelihood**
jj nn vbd rb   **8**   **3**   **9.02581**
(adjective-noun-verb,past tense-adverb)

M2S2
at an *American university meant very* differ   (RA22_14M2S2)
implicit *metalinguistic activity was about* 15   (RA30_C79 M2S2)
which *metalinguistic activity was also* impl   (RA30_4 M2S2)
implicit *metalinguistic activity was more* fre   (RA30_19 M2S2)
of the *statistical relationship were relatively* uncontroversial   (RA28_14 M2S2 + M3S1A)
*over the *traditional treatment was not* signifi   (RA7_16 M2S2)
*whereas the *traditional treatment was not* (mean   (RA7_17 M2S2)
M2S3
*quality of the *final output was not* clearly suppo   (RA16_C55 M2S3)

4-PoS tag Bundle in M2   **f**   **Expected**   **Log Likelihood**
vbd rb jj in   **7**   **2**   **8.52115**
(verb, past tense-adverb-adjective- preposition or subordinating conjunction)

M2S2
*activity *was also implicit.*   (RA30_4 M2S2)
it *was not possible from* the data   (RA28_30 M2S2)
it *was not rare for* him to   (RA19_37 M2S2)
difference *was not significant at* the usual   (RA7_2 M2S2)
*treatment *was not significant at* the 95%   (RA7_16 M2S2)
students *were also interested in* using the   (RA10_3 M2S2)
M2S3
there *were comparatively few of* the info   (RA9_7 M2S3)

4-PoS tag Bundle in M2   **f**   **Expected**   **Log Likelihood**
vbd in jj nns   **7**   **2**   **8.05925**
(verb, past tense- preposition or subordinating conjunction-adjective-noun plural)

M2S2
informant *commented that many Statistics* students   (RA28_70 M2S2)
they *consulted from academic journals* to   (RA21_29 M2S2)
He *emphasised that authentic data* rarel   (RA28_4 M2S2)
study*found that overall recasts* led   (RA3_54 M2S2)
student *mentioned that numerous conversations* with   (RA10_30 M2S2)
elicitations *occurred in different forms* with   (RA3_11 M2S2)
competence *varied across disciplinary communities,*   (RA22_5 M2S2

4-PoS tag Bundle in M2   f   Expected   Log Likelihood

vbd jj nn in   6   1   7.67037

(verb, past tense- adjective- noun-preposition or subordinating conjunction)

M2S2

hedges *drew particular interest from* my   (RA28_23 M2S2)

*Analysis *identified integrative motivation as*   (RA10_42 M2S2)

*participants *made significant improvement on* the   (RA10_23 M2S2)

M2S3

study   *expressed considerable enthusiasm for* DDL   (RA7_C55 M2S3)

however, *showed significant improvement for* grammaticality (RA5_C54 M2S3)

*learners *showed significant improvement on* interpretation   (RA5_C53 M2S3)

4-PoS tag Bundle in M2   f   Expected   Log Likelihood

vbd dt jj nn   17   7   7.12018

(verb, past tense-determiner- adjective- noun)

M2S2

*also *confirmed the textual analysis* in   (RA23_4 M2S2)

then, *employed the rhetorical conventions*   (RA24_C58 M2S2)

they *followed a native speaker* bigram   (RA1_2 M2S2)

learners *found the corpus data* easier to   (RA7_35 M2S2)

*already *had a sufficient level* of linguist   (RA22_17 M2S2)

experiment *had no measurable impact* on compre   (RA1_8 M2S2)

*analysis *identified a significant relationship*   (RA10_46 M2S2)

*who *invented a high number* of solution   (RA8_C27 M2S2)

that *reformulated the erroneous part* only   (RA3_12 M2S2)

*item *revealed a significant variation* in   (RA10_16M2S2)

vocabulary *made a significant contribution* to   (RA10_46 M2S2)

reading *made a significant contribution* to   (RA11_19 M2S2)

…suggested a varied range of   (RA21_31 M2S2)

writers *used a bare assertion* here , and   (RA28_22 M2S2)

 this *was a high stakes* decision : on   (RA28_31 M2S2)

M2S3

*program *had a positive impact on* listening   (RA4_2 M2S3)

*CF *had a positive effect* on the   (RA15_2 M2S3)

4-PoS tag Bundle in M2   f   Expected   Log Likelihood

in dt jj nn   93   68   6.08524

(preposition or subordinating conjunction-determiner-adjective-noun)

M2S1

…by the insignificant difference between   (RA18_3 M2S1)

…in the experimental group would sho   (RA6_23 M2S1 + M1S1)

…in the experimental group would rep   (RA6_28 M2S1 + M1S1)

...on the final test of listening   (RA6_2 M2S1 + M1S1)

M2S2

...about any potential patterns This   (RA28_35 M2S2)
..about the possible meaning of a   (RA28_8 M2S2)
...about the precise value to suggest   (RA28_24 M2S2)
*...*about the subject matter* , the mo   (RA25_26 M2S2)
...among the content categories My   (RA28_41 M2S2)
...as a mathematical subject and want   (RA28_70 M2S2)
*...*as a shy person* , did not   (RA25_30 M2S2)
...as an analytical tool to highlight   (RA24_C41 M2S2)
...as a significant predictor of stude   (RA10_42 M2S2)
...as the first variable to enter   (RA11_19 M2S2)
...as the first variable to enter   (RA11_28 M2S2)
...at a similar rate to his   (RA16_17 M2S2)
...at an international conference on '   (RA21_31M2S2)
*...*at an American university* meant   (RA22_14 M2S2)
*...*at the same time* , expressed the   (RA10_7 M2S2)
*...*At the same time* , though , ove   (RA21_35 M2S2)
...at the usual levels .   (RA7_2 M2S2)
...by the low percentage   (RA23_5 M2S2)
...by the local teachers " Cases   (RA36_21 M2S2)
*...*by the second recall* , it had   (RA2_6 M2S2)
*...*for the traditional treatment* , and   (RA7_18 M2S2)
...for the traditional treatment The q   (RA7_30 M3S1A + M2S2)
...for a significant reading variance   (RA11_47 M2S2)
...From the final texts , and from   (RA16_13M2S2)
...from the total number of typed   (RA16_13 M2S2)
...in a global job market   (RA10_6 M2S2)
...in a friendlier tone One thing   (RA23_4 M2S2)
*...*in a confirmatory tone* led to   (RA3_37 M2S2)
...in any subsequent calculations The   (RA28_31 M2S2)
...in the delayed recall test The   (RA2_7 M2S2)
*...*in the first year* to 33.5   (RA16_12 M2S2)
...in the instructional corpus was th   (RA9_4 M2S2)
...in the pervasive absence of Move   (RA24_7 M2S2)
...in the pragmatic benefits of Spani   (RA10_4 M2S2)
*...*in the present study* , as it   (RA2_7 M2S2)
*...*In the present study* , 70 %   (RA24_10 M2S2)
*...*In the present study* , for instance   (RA11_61M2S2)
*...*in the present study*, aspects o   (RA8_C32 M2S2)
**In the present study* , listening   (RA11_14 M2S2)
**In the present study* , the rate   (RA3_21 M2S2)
...in the previous example.   (RA28_21 M2S2)
*...*in the second year* , although his   (RA16_14 M2S2)
...in the second recall However ,   (RA2_5 M2S2)
**In the third year* , the quality   (RA16_6 M2S2)
**In this particular study* , implicit   (RA30_C79 M2S2)
...on a lexical level are spread   (RA1_40 M2S2)
...on either literal understanding or   (RA18_21 M2S2)
...of the academic communities despite   (RA22_43 M2S2)

...of the collaborative stance among	(RA23_3 M2S2)
...of the intact syllables of the	(RA9_11 M2S2)
...of the phonetic environment of the	(RA9_12 M2S3)
...of the prototypical IMRD paper	(RA19_1 M2S2)
...of the possessive determiner your	(RA9_23 M2S2)
...of the progressive verbs was less	(RA9_10 M2S2)
...of the progressive forms appeared	(RA9_11 M2S2)
...of the same feedback showed substa	(RA3_36 M2S2)
*...of the statistical relationship were	(RA28_14 M2S2 + M3S1A)
...of the various components	(RA8_9 M2S2)
...of the appropriate role of the	(RA26_19 M2S2)
*...of a statistical relationship , and	(RA28_7 M2S2)
*...of a statistical relationship , for	(RA28_8 M2S2)
*...over the traditional treatment was	(RA7_16 M2S2)
...on the Brazilian context . Besides	(RA24_32 M2S2)
...on a single handout The main	(RA7_C78 M2S2)
*...on the other hand , the superviso	(RA19_34 M2S2)
*On the other hand , the ratio	(RA24_38 M2S2)
*...that any particular value was an	(RA28_30 M2S2)
...that both morphological awareness vari	(RA11_21 M2S2)
...than the dictionary information	(RA7_35 M2S2)
...that the past subjunctive with inde	(RA5_38 M6S2A, dep. M2S2)
...than the perceptual salience of the	(RA9_11 M2S2)
...than the traditional treatment.	(RA7_36 M2S2)
...through a bare assertion .	(RA28_11 M2S2)
...through the extra work and challeng	(RA22_44M2S2)
...toward a differential effect of ins	(RA13_14M6S1A, dep. M2S2)
*...whereas the traditional treatment was	(RA7_17 M2S2)
...with the intrinsic motivation subscale	(RA14_13 M2S2)
*...within the Chinese program , the	(RA22_12M2S2)
...with the former accounting for more	(RA3_13 M2S2)
...with the practical application of r	(RA21_4 M3S1A + M2S2)
...with the Spanish language outside	(RA10_46 M2S2)
...with the Spanish language and	(RA10_43 M2S2)
...within a disciplinary community Hou	(RA22_19 M2S2)

M2S3

...for the apparent superiority of	(RA21_2 M2S3)
...in the second administration of the	(RA15_3 M2S3)
...of the current study found that	(RA5_C53 M2S3)
*...of the final output was not	(RA16_C55 M2S3)
...of the 3-year L2 study	(RA16_C55 M2S3)
*...that the different content categories	(RA28_9 M2S3)

---

*Part of them was captured in other PoS bundles.

O.   Formulaic Bundles Identified in Move 3

5-PoS tag Bundle in M3	f	Expected	Log Likelihood
nnp , cd : nnp (Proper noun, singular- ,- Cardinal number- :- Proper noun, singular)	**96**	**68**	**14.20561**

Ozeki , 2000 ; Seo , 2000   (RA4_12 M3S1A)

nnp cc nnp , cd (Proper noun, singular-coordinating conjunction-proper noun, singular- ,- cardinal number)	**76**	**55**	**9.73005**

Kroll & Stewart , 1994   (RA17_23 M3S3B)

cc nnp , cd : (coordinating conjunction-proper noun, singular-,-cardinal number- :)	**43**	**28**	**9.01037**

& Varela , 1998 ;   (RA3_C69 M3S1B)

cd : nnp cc (,- cardinal number-:-proper noun, singular- coordinating conjunction)	**47**	**31**	**8.85473**

, 2006; Joshi & Aaron   (RA11_4 M3S3A)

cd : nnp cc nnp (cardinal number- : - proper noun, singular- coordinating conjunction-proper noun, singular)	**48**	**32**	**8.56424**

2005; Johnston & Kirby   (RA11_4 M3S3A)

: nnp cc nnp , (: - proper noun, singular- coordinating conjunction- proper noun, singular- ,)	**46**	**31**	**8.09286**

; Collentine & Freed , 200   (RA10_18 M3S1B)

4-PoS tag Bundle in M3	f	Expected	Log Likelihood
vb vbn to dt (verb, base form- verb, past participle- to- determiner)	**12**	**6**	**6.41133**

M3S1
might indeed ***be justified to a*** greater   (RA19_23 M3S1B)
can not ***be limited to the*** act   (RA26_C38M3S1A)
M3S2
can ***be attributed to the*** small   (RA8_23 M3S2A)
may ***be attributed to the***   (RA11_22 M3S2A)

can not **be attributed to a**   (RA7_1 M2S2 + M3S2A)
could **be attributed to the** more   (RA3_48 M3S2A)
can **be related to the** more   (RA3_40 M3S2A)
have been found not to **be related to this** measure   (RA8_16 M3S2A)
might **be related to the** fact that   (RA23_15M3S2A)
might rather **be related to some** other   (RA8_20 M3S2B)
may **have contributed to this** distinction   (RA29_22 M3S2A)
M3S3
could **have contributed to the**   (RA11_35 M3S3C)

P.   Formulaic Bundles Identified in Move 4

5-PoS tag Bundle in M4	f	Expected	Log Likelihood
dt jj nn vbz dt	**5**	**0**	**10.65631**

M4S1
The present study is an endeavor   (RA17_C77 M4S1)
The present study underlines the complex   (RA18_C29 M4S1)
The present paper reports an analysis   (RA32_C26 M4S1)
M4S2
The present investigation supports the value   (RA10_57 M4S2)
with *the Spanish language has a* significant   (RA10_C96 M4S2)

5-PoS tag Bundle in M4	f	Expected	Log Likelihood
vbd vbn in dt nn	**5**	**0**	**7.23853**

M4S1
awareness *was adapted for this study*.   (RA11_C66 M4S1)
repair *was used as a measure* of   (RA3_C81 M4S1)
study *was designed with these points* in   (RA15_C63 M4S1)
tests *were used under the assumption* that   (RA11_C68 M4S1)
M4S2
vocabulary *was facilitated with the use* of   (RA12_C57 M4S2)

4-PoS tag Bundle in M4	f	Expected	Log Likelihood
dt jj nn vbz	**7**	**2**	**10.64000**

M4S1
*The present paper reports an analy   (RA32_C26 M4S1)
*The present study is an   (RA17_C77 M4S1)
*The present study underlines the   (RA18_C29 M4S1)
M4S2
that *the experimental procedure is* likely   (RA19_8 M4S2)
*The present investigation supports the   (RA10_57 M4S2)
, *the present study has* endeavoured   (RA1_C69 M4S1 + M4S2)
sharing *the same language is* not   (RA19_8 M4S2)

---

*Part of them was captured in other PoS bundles.

Q.   Formulaic Bundles Identified in Move 5

5-PoS tag Bundle in M5	**f**	**Expected**	**Log Likelihood**

rb vbn in dt nn   **6**   **1**   **6.39558**
(adverb- verb, past participle- preposition or subordinating conjunction-determiner-noun)

M5S1
was ***not included in the reaction*** time   (RA17_74 M5S1)
students had ***probably guessed that some sort*** of   (RA2_23 M5S1)
becoming ***rather bored with the routine*** The   (RA6_C64 M5S1)
M5S3
is ***apparently compensated by the provision***of   (RA7_21 M5S3)
was ***not designed with that purpose*** in   (RA1_9 M5S3)
was ***not practiced during the strategy*** instruction   (RA4_41 M5S3)

5-PoS tag Bundle in M5	**f**	**Expected**	**Log Likelihood**
dt nns in dt nn	**9**	**3**	**6.09378**

(determiner-noun- preposition or subordinating conjunction-determiner-noun)

M5S1
drawing ***any conclusions about the nature*** of   (A3_C77 M6S1A, (dep. M5S1)
of ***the findings from this study***.   (RA32_1 M5S1)
of ***the reviewers of this article*** pointed   (RA8_C35 M5S1)
based on ***the results of the study*** may   (RA20_29 M5S1)
Fourth, ***the students in the treatment*** groups   (RA15_49 M5S1)
M5S2
clarify ***the effects of this approach*** to   (RA6_5 M5S2)

5-PoS tag Bundle in M5	**f**	**Expected**	**Log Likelihood**
nn in dt nn vbz	**5**	**1**	**5.88242**

(noun- preposition or subordinating conjunction-determiner-noun-verb, present)

M5S1
another ***limitation of the study is*** that   (RA5_48 M5S1)
the ***point of the paper is*** not   (RA20_C36 M4S2, dep. M5S1)
M5S2
Another ***significance of this research lies*** in   (RA23_35 M5S2)
The ***significance of this study is*** threefold   (RA23_28 M5S2)
M5S3
The ***flexibility of the tool enables*** the   (RA16_C50 M5S3)
Rather, ***the findings of the study*** underscore   (RA22_C50 M5S2)

M5S3
as *the ethics of the study*.　　　　　　　　　　　　　　(RA20_34 M5S3)
of *the limitations of the study* to　　　　　　　　　　(RA20_34 M5S3)

4-PoS tag Bundle in M5	f	Expected	Log Likelihood
dt nns in dt	13	8	6.62690

(determiner-noun, plural- preposition or subordinating conjunction-determiner)

M5S1
of *the assessors in this* particular　　　　　　　　　(RA32_C30 M5S1)
creating *some overemphasis on those* items　　　　(RA2_22 M5S1)
*of *the findings from this* study　　　　　　　　　　(RA32_1 M5S1)
interpreting t*he implications of the* results　　　　(RA3_C75 M5S1)
of *the limitations of the* current　　　　　　　(RA13_12 M4S2 + M5S1)
*on *the results of the* study may　　　　　　　　　　(RA20_29 M5S1)
*of *the reviewers of this* article　　　　　　　　　　(RA8_C35 M5S1)
*Fourth , *the students in the* treatment　　　　　　(RA15_49 M5S1)
M5S2
*clarity *the effects of this* approach to　　　　　　(RA6_5 M5S2)
*Rather, *the findings of the* study undersco　　　(RA22_C50 M5S2)
addressed *the needs of a* specific learner　　　　(RA4_C64 M5S2)
M5S3
minimize *the effects of the* limitations of　　　　(RA20_34 M5S3)
*well as *the ethics of the* study　　　　　　　　　　(RA20_34 M5S3)

4-PoS tag Bundle in M5	f	Expected	Log Likelihood
vbn in dt nn	21	10	5.98866

(Verb, past participle- preposition or subordinating conjunction-determiner-noun)

M5S1
*rather *bored with the routine* The learner　　　　(RA6_C64 M5S1)
*were *displayed on the top* of the　　　　　　　　(RA2_22 M5S1)
*were *excluded from the analyses.*　　　　　　　　(RA1_31 M5S1)
*probably *guessed that some sort* of recall　　　(RA2_23 M5S1)
was *included in this meta‑analysis* Howeve　　(RA12_C49 M5S1)
*not *included in the reaction* time anal　　　　　(RA17_74 M5S1)
*were *involved in the use* of our　　　　　　　　(RA2_16 M5S1)
been *noted that this experiment* provides　　　　(RA1_33 M5S1)
*generalization *based on the results* of the　　　(RA20_29 M5S1)
however , *limited in the scale* and scope　　　　(RA32_C28 M5S1)
findings *reported in this article*, of　　　　　　　(RA35_C35 M5S1)
those *reported in this study*, we　　　　　(RA19_12 M3S1A, (dep. M5S1)

M5S2

is **removed from the context** of produc	(RA16_C59 M5S2)
research **reported in this article** is a	(RA15_C68 M5S2)
be **seen as a strength** of the	(RA4_51 M5S2)
*was **targeted for the provision** of CF	(RA15_4 M5S2)

M5S3

*apparently **compensated by the provision** of si	(RA7_21 M5S3)
*not **designed with that purpose** in mind	(RA1_9 M5S3)
*not **practiced during the strategy** instruc	(RA4_41 M5S3)
As **illustrated through the findings** pre	(RA16_C54 M5S3)
findings **presented in this article**, this	(RA16_C54 M5S3)

---

*Part of them was captured in other PoS bundles.

R.   Formulaic Bundles Identified in Move 6

5-PoS tag Bundle in M6	**f**	**Expected**	**Log Likelihood**
md vb jj to vb	**15**	**7**	**9.60619**

(modal- verb, base form-adjective-to-verb, base form)

M6S1
it *may be necessary to examine* the   (RA13_16 M6S1B)
it *might be hard to examine* all   (RA3_C98 M6S1B)
differences *will be important to quantify* and   (RA9_50 M6S1B)
it *might be worthwhile to analyze* the   (RA8_C38 M6S1A)
it *would be important to examine* not   (RA22_C59 M6S1A)
it *would be interesting to examine* the   (RA18_C43M6S1A)
it *would be interesting to see* if   (RA2_9 M6S1B)
it *would be interesting to learn* how   (RA24_C70 M6S1A)
it *would be interesting to understand* why   (RA24_C68 M6S1A)
participants *would be necessary to establish* with   (RA8_C34 M6S1B)
it *would be worthwhile to proceed* along   (RA19_C52M6S1A)
M6S2
thus *may be able to draw* learners '   (RA17_48 M6S2A)
disciplines *may be able to pass* on   (RA26_C55 M6S2A)
They *would be able to predict* problemati   (RA17_48 M6S2A)
area *would be unable to do*.   (RA26_C55 M6S2A)

5-PoS tag Bundle in M6	**f**	**Expected**	**Log Likelihood**
md vb vbn to vb	**13**	**5**	**9.58003**

(modal-verb, base form-verb, past participle-to- verb, base form)

M6S1
what *can be done to promote* its   (RA7_54 M6S1A)
instruction *may be required to oust* a   (RA5_30 M6S1B)
input *may be enhanced to increase* the   (RA9_52 M6S1B)
questionaire *should be administered to ask* about   (RA10_82 M6S1B)
tasks *should be explored to determine* whether   (RA12_C58 M6S1A)
M6S2
insiders, *can be asked to consider* both   (RA28_69 M6S2A)
tasks *can be devised to focus* on   (RA9_C68 M6S2A)
they *could be instructed to contact* a   (RA26_C50 M6S2A)
students *could be found to help* their   (RA33_C56 M6S2A)
data *should be collected to identify* teachers'   (RA35_30 M6S2A)
students *should be encouraged to use* the   (RA33_5 M6S2A)
teachers *should be encouraged to observe* pilot   (RA35_27 M6S2A)
proofreaders *should be permitted to work* on   (RA26_C54 M6S2A)

5-PoS tag Bundle in M6	f	Expected	Log Likelihood
nn, nns md vb	**11**	**4**	**8.84394**
(noun- noun, plural-modal-verb, base form)			

M6S2

In ***addition, instructors can utilize*** specific	(RA25_C55 M6S2A)
student ***behavior, teachers can promote*** students'	(RA14_24M6S2B)
of ***competence, teachers should use*** individual	(RA14_29M6S2B)
linguistic ***correctness, teachers may achieve*** better	(RA15_C77M6S2A)
task-based ***course, departments should provide*** multiple	(RA35_25 M6S2A)
abroad ***environment , researchers might begin*** to	(RA10_C106 M6S1A)
language ***instruction, learners can benefit*** from	(RA12_10 M6S2A)
to ***motivation , instructors should attempt*** to	(RA10_60 M6S2A)
During ***orientation, participants should be*** provided	(RA10_79 M6S2A)
inferential ***reading, testers should be*** sensitive	(RA18_C34 M6S2A)
same ***time, instructors can provide*** students	(RA10_64 M6S2A)

5-PoS tag Bundle in M6	f	Expected	Log Likelihood
nns md vb to vb	**10**	**4**	**6.37966**
(noun, plural-modal-verb, base form- to- verb, base form)			

M6S1

L2 ***learners will need to be*** carefully	(RA17_76 M6S1B)
approaches, ***researchers may need to examine*** different	(RA13_C23 M6S1A)
*environment , ***researchers might begin to investigate*** how	(RA10_C106 M6S1A)
some ***researchers might choose to write*** their	(RA24_C68 M6S1A)

M6S2

*motivation , ***instructors should attempt to incorporate*** activities	(RA10_60 M6S2A)
prepared ***materials would seem to provide*** one	(RA7_C62M6S2B)
cases, ***teachers can try to understand*** why learners	(RA17_50 M6S2A)
language ***teachers may resort to writing*** tasks	(RA2_15 M6S2A)
English ***teachers might seek to increase*** students'	(RA14_C47M6S2A)
English ***teachers should strive to provide*** such	(RA14_26M6S2A)

5-PoS tag Bundle in M6	f	Expected	Log Likelihood
md vb dt nn in	**15**	**8**	**5.97519**
(modal-verb, base form-determiner-noun- preposition or subordinating conjunction)			

M6S1

factors that ***can affect the usefulness of*** feedback	(RA3_C97M6S1A)
and ***could help the selection of*** language	(RA8_C31M6S1A)
enquiry ***could encompass an examination of*** the	(RA18_C45 M6S1A)
factors ***may outweigh the influence of*** examinations	(RA31_24 M6S1A)
this ***might be an avenue for*** future	(RA4_56 M6S1A)
future research ***might probe this issue with*** different	(RA18_C42 M6S1B)
section) ***will be the subject of*** future	(RA1_C58 M6S1A)

M6S2

that ***can increase the amount of*** input	(RA17_63M6S2A)
logging ***can provide a means for*** the	(RA16_39 M6S2A)
classroom, ***could become the basis for*** a	(RA36_46 M6S2A)
current study ***may include an identification of*** a	(RA32_6 M6S2A)
learning ***will foster a sense of*** ownership	(RA14_22 M6S2B)
institution ***should consider the effectiveness of*** the	(RA33_C51 M6S2A)
users ***should develop an awareness of*** the	(RA32_24 M6S2B)
we ***should discuss the relationship between*** form	(RA19_C45 M6S2A)

5-PoS tag Bundle in M6	f	Expected	Log Likelihood
prp md vb jj to	**10**	**4**	**5.70554**
(Personal pronoun-modal-verb, base form-adjective-to)			

M6S1

*learners ***it may be necessary to*** examine	(RA13_16 M6S1B)
*It might be hard to examine	(RA3_C98 M6S1B)
*creativity, ***it might be worthwhile to*** analyze	(RA8_C38 M6S1A)
First, ***it would be interesting to*** understand	(RA24_C68M6S1A)
*Casanave (1998) ***it would be interesting to*** lear**n**	(RA24_C70M6S1A)
example, ***it would be interesting to*** examine	(RA18_C43 M6S1A)
*study, ***it would be interesting*** to see	(RA2_9 M6S1B)
research, ***it would be important to*** examine	(RA22_C59M6S1A)
think ***it would be worthwhile to*** proceed	(RA19_C52 M6S1A)
M6S2	
*They would be able to predict	(RA17_48 M6S2A)

5-PoS tag Bundle in M6	f	Expected	Log Likelihood
vb dt nn in dt	**17**	**10**	**5.70014**
(verb, base form-determiner-noun-preposition or subordinating conjunction-determiner)			

M6S1

needed to ***examine the effectiveness of these*** instructional	(RA13_17 M6S1A)
needed to ***examine the effectiveness of the*** selective	(RA15_56 M6S1A)
could ***encompass an examination of the*** process	(RA18_C45M6S1A)
is important to ***determine the applicability of the*** results	(RA14_C40 M6S1B)
would both ***improve the power of the*** statistical	(RA16_32 M6S1B)
research to ***determine the nature of the*** knowledge	(RA30_C88 M6S1A)
and to ***explore the role of this*** activity	(RA30_C88 M6S1A)
to ***ascertain the level of the*** representations	(RA30_74 M6S1B)
M6S2	
*can ***provide a means for the*** writers	(RA16_39 M6S2A)
tend to ***have a background in the*** humanities	(RA19_C42 M6S2A)
may also ***be a need for the*** teachers	(RA20_17 M6S2A)
*may ***include an identification of a*** critical	(RA32_6 M6S2A)

*should *consider the effectiveness of the* place   (RA33_C51 M6S2A)
textbook to *dictate the content of the* course   (RA35_23 M6S2A)
and *discuss the effectiveness of those* routines   (RA35_32 M6S2A)
*could *become the basis for a* form   (RA36_46 M6S2A)
*should *develop an awareness of the* situate   (RA32_24 M6S2B)

5-PoS tag Bundle in M6	f	Expected	Log Likelihood
nn md vb dt nn	**13**	**5**	**5.50200**

(noun- modal- verb, base form-determiner-noun)

M6S1
*of *inquiry could encompass an examination* of   (RA18_C45 M6S1A)
*Future *research might probe this issue* with   (RA18_C42 M6S1B)
future *research might use an eye tracking* methodology   (RA1_54 M6S1B)
Further *research should explore the beliefs*, attitudes   (RA36_C53 M6S1A)
further *work might increase the learners '* confidence   (RA7_50 M5S1 + M6S1B)
M6S2
task-based *approach could include the following*: grammar (RA31_5 M6S2A)
and *connection will benefit all students*, local   (RA25_C63 M6S2A)
writing *event can stimulate the writer* to   (RA16_40 M6S2B)
the *institution should consider the effectiveness* of   (RA33_C51 M6S2A)
*of *learning will foster a sense* of   (RA14_22 M6S2B)
*current *study may include an identification* of   (RA32_6 M6S2A)
of *study may provide a way* forward   (RA21_37 M6S2A)
the *team could provide a forum* that   (RA36_42 M6S2A)

5-PoS tag Bundle in M6	f	Expected	Log Likelihood
, prp md vb jj	**7**	**4**	**5.50080**

(,-personal pronoun-modal-verb, base form-adjective)

M6S1
creativity, it might be worthwhile* to   (RA8_C38 M6S1A)
research*, it would be important* to   (RA22_C59 M6S1A)
For example, it would be interesting* to   (RA18_C43M6S1A)
study, it would be interesting* to   (RA2_9 M6S1B)
First, it would be interesting* to   (RA24_C68 M6S1A)
M6S2
Therefore, *we should provide multiple* opportunities for   (RA14_31 M6S2A)
performance*, it would be helpful for* such classes   (RA25_C51 M6S2A)

5-PoS tag Bundle in M6	**f**	**Expected**	**Log Likelihood**
rb, jj nns md	**6**	**2**	**5.09202**

(adverb- adjective-noun, plural- modal)

M6S1

Furthermore, future studies should compare   (RA10_C104 M6S1A)
Moreover, further studies could also   (RA8_C37 M6S1A)
Therefore, future studies should examine   (RA12_C56 M6S1A)
Therefore, vocabulary researchers should use   (RA12_C38 M6S1B)
M6S2
Moreover, social activities could be   (RA33_C47 M6S2A)
Therefore, English teachers should   (RA14_26 M6S2A)

5-PoS tag Bundle in M6	**f**	**Expected**	**Log Likelihood**
nns md vb vbn in	**11**	**5**	**5.95336**

(noun, plural-modal-verb, base form-verb, past participle- preposition or subordinating conjunction)

M6S1

more *data should be collected in* both   (RA3_C95 M6S1B)
word *factors should be taken into* consideration   (RA2_12 M6S2A + M6S1A)
participating *teachers can be recruited from* the   (RA34_C32 M6S1B)
of s*takeholders can be promoted in* Hong Kong   (RA36_C54 M6S1A)
M6S2
those *concerns may be addressed by* revising   (RA35_31 M6S2A)
and *decisions can be filtered through* their   (RA20_C37 M6S2A)
and *exams could be reinforced if* tasks   (RA31_21 M6S2A)
repertoire *goals could be addressed by* engaging   (RA28_61 M6S2A)
Learners should be provided with information   (RA35_28 M6S2A)
orientation, *participants should be provided with* social   (RA10_79 M6S2A)
reading *topics should be selected with* a   (RA18_C36 M6S2A)

4-PoS tag Bundle in M6	**f**	**Expected**	**Log Likelihood**
nn nns md vb	**18**	**9**	**8.66302**

(noun- noun, plural-modal-verb, base form)

M6S1

at which *point learners may become* sensitive   (RA13_C23 M6S1A)
*thet *word factors should be* taken into   (RA2_12 M6S2A + M6S1A)
unrehearsed *student observations might provide* important   (RA10_C102 M6S1B)
multiple *morphology measures would be* an appropriate   (RA11_C69 M6S1B)
proficiency *L2 learners will need* to be   (RA17_76 M6S1B)
*suggests, *teacher factors may outweigh* the   (RA31_24 M6S1A)
or *training programs can help* L2   (RA35_C42 M6S1A)

M6S2

proficient *L2 students can be* an	(RA2_13 M6S2A)
*that *language teachers may resort* to writing	(RA2_15 M6S2A)
input, *corpus tools can hel*p us extract	(RA17_64 M6S2A)
inferential *comprehension questions should be* carefully	(RA18_C37M6S2A)
located, *heritage speakers may benefit* more	(RA5_38 M6S2A, dep. M2S2)
*language *repertoire goals could be* addressed	(RA28_61 M6S2A)
that *student teachers should adopt* these	(RA32_23 M6S2A)
***Learners should be* provided	(RA35_28 M6S2A)
what *extent universities would permit* such	(RA26_12 M6S2B)
such *discourse norms could be* through	(RA32_25 M6S2B)
one/*ones can be copied* , which	(RA33_C43M6S2B)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
jj nns md vb	**37**	**22**	**12.53143**
(adjective-noun, plural-modal- verb, base form)			

M6S1

Additional studies will help clarify	(RA12_C53 M6S1A)
from *Brazilian researchers would help* explore	(RA24_C51 M6S1B)
of *Brazilian scholars would cast* light	(RA24_C67 M6S1B)
some *brief remarks can be* made.	(RA26_C42 M6S1A)
*these *different stakeholders can be* promote	(RA36_C54 M6S1A)
*In addition , *future studies should compare* the	(RA10_C104 M6S1A)
Therefore ,*future studies should examine* in a	(RA3_C89 M6S1B)
*Furthermore , *future studies should examine* the	(RA12_C56 M6S1A)
that *future studies will find* a way	(RA2_19 M6S1B)
the *following areas can help* L2	(RA35_C37 M6S1A)
what *instructional methods may be* best	(RA5_39 M6S1A)
and *subsequent revisions can be* found	(RA23_C64 M6S1A)
Such data would be useful in	(RA1_55 M6S1B)
Such analyses should be sensitive	(RA9_37 M6S1B)
Whether *verbal interactions would reveal* different	(RA23_C63 M6S1A)
*Therefore ,*vocabulary researchers should use* a	(RA12_C38 M6S1B)

M6S2

age - *appropriate narratives may be* a particularly	(RA9_55 M6S2A)
that *Brazilian researchers might write* articles	(RA24_C61 M6S2A)
English *Brazilian authors might be* in a	(RA24_C62 M6S2A)
in *elementary schools might be* difficult	(RA14_26 M6S2A)
*motivation *English teachers might seek* to	(RA14_C47 M6S2A)
*difficult , *English teachers should strive* to	(RA14_26 M6S2A)
more *explicit links could be* built and	(RA31_16 M6S2A)
following *general suggestions may serve* as	(RA35_21 M6S2A)
that *individual instructors can do* much	(RA25_C52 M6S2A)
*If *instructional tasks can be* devised	(RA9_C68 M6S2A)
suggests that *L2 educators should pay* attention	(RA17_46 M6S2A, dep. M6S1A)
that *L2 teachers can draw* on	(RA35_C34 M5S2 + M6S2A)
the *natural sciences can be* problematic	(RA19_C41 M6S2A)

that *prepared materials can provide* and   (RA7_C68 M6S2B)
*using *prepared materials would seem* to   (RA7_C62 M6S2B)
producing *positive results might inspire* publis   (RA7_C81 M6S2A)
less *skilled listeners can benefit* the   (RA6_C48 M6S2A)
*Moreover , *social activities could be* introduced   (RA33_C47 M6S2A)
which *such policies must be* implemented   (RA20_15 M6S2A)
into *task-based courses may be* particularly   (RA35_11 M6S2A)
*their *own disciplines may be* able to   (RA26_C55 M6S2A)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
jj nn md vb	**34**	**22**	**7.89086**
(adjective- noun-modal-verb, base form)			

M6S1
Therefore, *future research should examine* how   (RA3_C94 M6S1A)
Such research will allow researchers   (RA11_C70 M6S1A)
Further research should investigate factors   (RA12_C55 M6S1A)
the *present finding could be* generalized   (RA17_C81 M6S1A)
the *present study could be* tested in   (RA18_C48 M6S1A)
above, *future research might use* an eye   (RA1_54   M6S1B)
conditions, *such research can contribute* significantly   (RA3_C80 M6S1B)
the *current study could be* further enriched   (RA6_C55 M6S1B)
gains, *further work might show* what differ   (RA7_C74 M6S1B)
incidental *vocabulary learning could include* language   (RA12_C61 M6S1B)
an *L1 composition would have* permitted   (RA16_33 M6S1B)
Future research might probe this issue   (RA18_C42 M6S1B)
*but *further work might increase* the lea   (RA7_50 M5S1 + M6S1B)
their *cultural capital would enjoy* a high   (RA22_C58 M6S1A)
of *empirical research will help* L2   (RA35_C47 M6S1A)
Further research should explore the   (RA36_C53 M6S1A)
contexts *such engagement will be* neither a   (RA21_C60 M6S1B)
an *ethnographic study might help* explain   (RA24_C67 M6S1B)
the *collaborative task would provide* information   (RA30_75 M6S1B)
M6S2
a *specific task can be* linked in   (RA4_48 M6S2A)
this *extra step might provide* a valuable   (RA17_56 M6S2A)
setting, *such input could include* aspects   (RA16_43 M6S2B)
which *such work can be* meaningfully commu   (RA21_23 M6S2A)
English *conventionalized genre may provide* L2   (RA27_C26 M6S2A)
improving *professional development might benefit* from   (RA29_71 M6S2A)
A *useful strategy might be* to focus   (RA31_33 M6S2A)
an *appropriate balance can be* found b   (RA31_C36 M6S2A)
*the *current study may include* an identi   (RA32_6 M6S2A)
involve *genre analysis would be* effective m   (RA32_17 M6S2A)
schools, *such problematization might be* achieved   (RA36_35 M6S2A)
this *critical consciousness might be* achieved   (RA36_43 M6S2A)
the *key distinction should be* not whether   (RA21_25 M6S2B)
academic *professional genre may be* translated   (RA27_C24 M6S2B)

the *reflective discourse would enable* student   (RA32_12 M6S2B)

4-PoS tag Bundle in M6   **f**   **Expected**   **Log Likelihood**
md vb to vb   **21**   **9**   **11.41107**
(modal- verb, base form-to-verb, base form)

M6S1
*researchers *may need to examine* different   (RA13_C23 M6S1A)
*researchers *might begin to investigate* how for   (RA10_C106 M6S1A)
researchers *might choose to write* their artic   (RA24_C68 M6S1A)
*learners *will need to be carefully* considered   (RA17_76 M6S1B)
writing *would continue to degrade* up to   (RA2_9   M6S1B)
L2 *would need to be* examined in   (RA17_C81 M6S1A)
position *would help to explore* these possibi   (RA16_28 M6S1A)

M6S2
they *can elect to focus* variably on   (RA15_C76 M6S2A)
*teachers *can try to understand* why learners   (RA17_50 M6S2A)
which *may need to be* edited by   (RA17_65 M6S2A)
*teachers *may resort to writing* tasks that   (RA2_15 M6S2A)
these *might help to convince* a wider   (RA7_C76 M6S2A)
*strategy *might be to focus* more on   (RA31_33 M6S2A)
*teachers *might seek to increase* students '   (RA14_C47 M6S2A)
teacher *must continue to learn* about teachi   (RA29_41 M6S2A)
This *would seem to be* an important   (RA19_C47 M6S2B)
here *would seem to support* this use   (RA7_C63 M6S2A)
*instructors *should attempt to incorporate* activities   (RA10_60 M6S2A)
therefore, *should continue to encourage* students   (RA10_89 M6S2A)
*materials *would seem to provide* one obvious   (RA7_C62 M6S2B)
*teachers *should strive to provide* such envir   (RA14_26 M6S2A)

4-PoS tag Bundle in M6   **f**   **Expected**   **Log Likelihood**
md vb dt nn   **35**   **16**   **10.78300**
(modal- verb, base form- determiner-noun)

M6S1
*that *can affect the usefulness* of feedba   (RA3_C97 M6S1A)
*inquiry *could encompass an examination* of   (RA18_C45 M6S1A)
and *could help the selection* of language   (RA8_C31 M6S1A)
*factors *may outweigh the influence* of exami   (RA31_24 M6S1A)
*This *might be an avenue* for future   (RA4_56 M6S1A)
*work *might increase the learners* ' confidence   (RA7_50 M5S1 + M6S1B)
*research *might probe this issue* with differe   (RA18_C42 M6S1B)
*research *might use an eye* tracking methodolo   (RA1_54   M6S1B)
*studies *should compare the roles* of motiva   (RA10_C104 M6S1A)
*studies *should examine the relationships* between   (RA10_C105 M6S1A)

*research *should explore the beliefs* , attitudes (RA36_C53 M6S1A)
*researchers *should use a research* design that (RA12_C38 M6S1B)
section *will be the subject* of future (RA1_C58 M6S1A)
*studies *will find a way* to measure (RA2_19   M6S1B)
which *would allow the students* to reflec (RA10_C100 M6S1B)

M6S2
we *can help the students* to develop (RA19_C43 M6S2A, dep. M6S2B)
that *can increase the amount* of input (RA17_63 M6S2A)
*logging *can provide a means* for the (RA16_39 M6S2A)
*event *can stimulate the writer* to reflect (RA16_40 M6S2B)
*classroom, *could become the basis* for a (RA36_46 M6S2A)
*approach *could include the following* : grammar (RA31_5 M6S2A)
*team *could provide a forum that* emphasizes (RA36_42 M6S2A)
study *may include an identification* of a (RA31_5 M6S2A)
which *might influence a writer* to make (RA28_69 M6S2A)
study *may provide a way* forward here , (RA21_37 M6S2A)
*institution *should consider the effectiveness* of (RA33_C51 M6S2A)
*teachers *should create a classroom* environment (RA14_33 M6S2B)
*users *should develop an awareness* of the (RA32_24 M6S2B)
we *should discuss the relationship* between (RA19_C45 M6S2A)
1992 *should provide an impetus* to ESL (RA23_C58 M6S2B)
*connection *will benefit all students* , local (RA25_C63 M6S2A)
*learning *will foster a sense* of ownership (RA14_22 M6S2B)
authorities *will wish the write*r's supervisor (RA26_C51 M6S2A)
universities *would have an overview* and (RA26_C45 M6S2A)
This *would represent a learning* opportunity (RA36_47 M6S2A)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
nn md vb nns	12	5	9.58003
(noun-modal-verb, base form, noun, plural)			

M6S1
this *context could include teachers* ' beliefs (RA20_31 M6S1A)
of research *engagement can deepen understandings* of (RA21_C60M6S1A)
such a *study could have implications* for (RA1_56 M6S1A)
*Such *research will allow researchers* to (RA11_C70 M6S1A)
*Further *research should investigate factors* regarding (RA12_C55 M6S1A)
This *work will inform efforts* to provide (RA29_C86 M6S1A)
M6S2
the *administration should find ways* to (RA33_C53M3S1A+M6S2A)
*such *input could include aspects* of English (RA16_43 M6S2B)
program *staff should help students* develop (RA10_85 M6S2A)
of *research would enable teachers* to (RA21_33 M6S2A)
The *institution should arrange workshops* where (RA33_22 M6S2A)
this *team could provide NETs* with legitimate (RA36_40 M6S2A)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
nns md vb jj	23	11	9.37204

(noun, plural-modal-verb, base form-adjective)

M6S1
*such ***analyses should be sensitive*** to   (RA9_37 M6S1B)
*Such ***data would be useful*** in gaining   (RA1_55 M6S1B)
*These ***differences will be important*** to   (RA9_50 M6S1B)
*verbal ***interactions would reveal different*** stances   (RA23_C63 M6S1A)
*which point ***learners may become sensitive*** to   (RA13_C23 M6S1A)
*student ***observations might provide important*** information   (RA10_C102 M6S1B)
of ***participants would be necessary*** to   (RA8_C34 M6S1B)

M6S2
Task - based ***approaches could be feasible*** in diverse   (RA31_C36 M6S2A)
course, ***departments should provide multiple*** opportunities   (RA35_25 M6S2A)
*own ***disciplines may be able*** to pass   (RA26_C55 M6S2A)
where ***guidelines would be beneficial*** stand   (RA26_C43 M6S2A)
discussions, ***instructors can supplement whole*** class   (RA25_C58 M6S2A, dep. M2S2)
and **instructors can do more** to help   (RA25_C45 M6S2A)
addition, ***instructors can utilize specific*** scaffolding   (RA25_C55 M6S2A)
extent ***learners may have few*** immediate or   (RA35_3 M6S2B)
*natural ***sciences can be problematic*** . This   (RA19_C41 M6S2A)
*elementary ***schools might be difficult***, English   (RA14_26 M6S2A)
correctness, ***teachers may achieve better*** results   (RA15_C77 M6S2A)
competence, ***teachers should use individual*** criteria   (RA14_29 M6S2B)
*reading , ***testers should be sensitive*** to the   (RA18_C34 M6S2A)
the ***topics can represent diverse*** cultures   (RA25_C57 M6S2A)
their ***products may obtain lower*** interest   (RA24_C54 M6S2B)
*extent ***universities would permit such*** exchanges   (RA26_12 M6S2B)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
vb vbn to vb	16	7	8.99694

(verb, base form- verb, past participle- to-verb, base form)

M6S1
teaching ***be adapted to optimize*** learning   (RA35_C39 M6S1A)
*should ***be administered to ask*** about   (RA10_82 M6S1B)
*what can ***be done to promote*** its use   (RA7_54 M6S1A)
*input may ***be enhanced to increase*** the   (RA9_52 M6S1B)
*tasks should ***be explored to determine*** whether   (RA12_C58 M6S1A)
*instruction may ***be required to oust*** a competing   (RA5_30 M6S1B)
M6S2
*can ***be asked to consider*** both the   (RA28_69 M6S2A)
*should ***be collected to identify*** teachers   (RA35_30 M6S2A)
*can ***be devised to focus*** on different   (RA9_C68 M6S2A)
*should ***be encouraged to observe*** pilot   (RA35_27 M6S2A)
*should ***be encouraged to use*** the SAC   (RA33_5 M6S2A)

teachers *be encouraged to transform* their   (RA34_C25 M6S2A)
could *be found to help* their friends   (RA33_C56 M6S2A)
*could *be instructed to contact* a univers   (RA26_C50 M6S2A)
*should *be permitted to work* on texts   (RA26_C54 M6S2A)
learners *be required to make* predictions ,   (RA6_C68 M6S2A)

4-PoS tag Bundle in M6	**f**	**Expected**	**Log Likelihood**
md vb jj to	**18**	**9**	**8.66302**

(modal-verb, base form-adjective-to)

M6S1
*learners *may become sensitive to* different   (RA13_C23 M6S1A)
*it *may be necessary to* examine   (RA13_16 M6S1B)
*It *might be hard to* examine   (RA3_C98 M6S1B)
*it *might be worthwhile to* analyze   (RA8_C38 M6S1A)
*analyses *should be sensitive to* discourse   (RA9_37 M6S1B)
*differences *will be important to* quantify   (RA9_50 M6S1B)
*it *would be important to* examine not   (RA22_C59 M6S1A)
*it *would be interesting to* examine   (RA18_C43 M6S1A)
*it *would be interesting to* understand   (RA24_C68 M6S1A)
*it *would be interesting to* learn how   (RA24_C70 M6S1A)
*it *would be interesting to* see   (RA2_9 M6S1B)
*participants *would be necessary to* establish   (RA8_C34 M6S1B)
it *would be worthwhile to* proceed along   (RA19_C52 M6S1A)
M6S2
*thus *may be able to* draw learners'   (RA17_48 M6S2A)
*disciplines *may be able to* pass on   (RA26_C55 M6S2A)
*testers *should be sensitive to* the type   (RA18_C34 M6S2A)
*They *would be able to* **predict** problematic   (RA17_48 M6S2A)
area *would be unable to* do   (RA26_C55 M6S2A)

4-PoS tag Bundle in M6	**f**	**Expected**	**Log Likelihood**
nn md vb dt	**22**	**12**	**8.58435**

(noun-modal-verb, base form-determiner)
M6S1
*cultural *capital would enjoy a* higher marke   (RA22_C58 M6S1A)
this *idea would be a* suitable topic   (RA11_64 M6S1A)
*of *inquiry could encompass an* examination   (RA18_C45 M6S1A)
the *input may require a* more explicit   (RA9_62 M6S1A)
of *reference may be the* Japanese conte   (RA31_24 M6S1A)
*Future *research might probe this* issue with   (RA18_C42 M6S1B)
*future *research might use an* eye tracking   (RA1_54 M6S1B)
*Further *research should explore the* beliefs   (RA36_C53 M6S1A)
*further *work might increase the* learners '   (RA7_50M5S1 + M6S1B)
M6S2
*task-based *approach could include the* following   (RA31_5 M6S2A)
task-based *approach may demand a* varied repert   (RA31_32 M6S2A)

*and *connection will benefit all* students (RA25_C63 M6S2A)
writing *event can stimulate the* writer to (RA16_40M6S2B)
*the *institution should consider the* effect (RA33_C51 M6S2A)
*of *learning will foster a* sense of (RA14_22 M6S2B)
CARS *model may be a* good strategy. (RA24_C53 M6S2A)
*extra *step might provide a* valuable oppor (RA17_56 M6S2A)
this *study can provide a* supporting tool (RA27_C24 M6S2B)
*current *study may include an* identification (RA32_6 M6S2A)
*of *study may provide a* way forward (RA21_37 M6S2A)
of *teaching may affect the* future deve (RA29_71 M6S2A)
the *team could provide a* forum that (RA36_40 M6S2A)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
nns md vb vbn	**30**	**19**	**8.56404**

(noun, plural-modal-verb, base form-verb, past participle)

M6S1
these *activities might be increased* or (RA9_51   M6S1B)
*more *data should be collected* in both (RA3_C95 M6S1B)
their *effects must be considered* when (RA3_C99 M6S1B)
*word *factors should be taken* into cons (RA2_12 M6S2A + M6S1A)
*brief *remarks can be made* . (RA26_C42 M6S1A)
*subsequent *revisions can be found* . (RA23_C64 M6S1A)
*different *stakeholders can be promoted* in (RA36_C54 M6S1A)
*these *tasks should be explored* to deter (RA12_C58 M6S1A)
between *teachers can be kept* as constant (RA34_C32 M6S1B)
*participating *teachers can be recruited* from th (RA34_C32 M6S1B)
these *variables should be kept* the same (RA34_C30 M6S1B)
M6S2
*social *activities could be introduced* more (RA33_C47 M6S2A)
*those *concerns may be addressed* by revi (RA35_31 M6S2A)
*of *data should be collected* to ident (RA35_30 M6S2A)
*and *decisions can be filtered t*hrough (RA20_C37 M6S2A)
*and *exams could be reinforced* if task (RA31_21 M6S2A)
*repertoire *goals could be addressed* by engag (RA28_61 M6S2A)
***Learners should be provided* with (RA35_28 M6S2A)
explicit *links could be built* and articula (RA31_16 M6S2A)
*one/*ones can be copied*, which can (RA33_C43 M6S2B)
*orientation, *participants should be provided* with (RA10_79 M6S2A)
*such *policies must be implemented* needs (RA20_15 M6S2A)
*whether *proofreaders should be permitted* to (RA26_C54 M6S2A)
valued *practices could be enhanced*. (RA32_21 M6S2A)
some *students could be found* to help (RA33_C56 M6S2A)
*that *students should be encouraged* to (RA33_5 M6S2A)
*instructional *tasks can be devised* to focus (RA9_C68 M6S2A)
*possible, *teachers should be encouraged* to (RA35_27 M6S2A)
*reading *topics should be selected* with a (RA18_C36 M6S2A)
These *voices should be heeded* and taken (RA23_42 M6S2A)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
vb nns to vb	**14**	**7**	**7.47171**
(verb, base form, noun, plural-to-verb, base form)			

M6S1

would ***allow listeners to compare*** concatenat   (RA6_C57 M6S1B)
will ***help researchers to garner*** more r   (RA11_C74 M6S1B)
*will ***inform efforts to provide*** professional   (RA29_C86 M6S1A)
would ***assist teachers to have*** a more   (RA34_22 M6S1B)
M6S2
*might ***inspire publishers to produce*** DDL   (RA7_C81 M6S2A)
helps ***prepare students to deal*** with the   (RA9_53 M6S2A)
*to ***encourage students to live*** with ho   (RA10_89 M6S2A)
classroom ***enable students to choose*** the obje   (RA14_25 M6S2B)
would ***enable teachers to avoid*** overly br   (RA21_33 M6S2A)
to ***train students to become*** successful   (RA23_42 M6S2A)
to ***help learners to plan*** their stud   (RA33_C45 M6S2A)
not ***take opportunities to focus*** on the   (RA35_18 M6S2A)
should ***find ways to increase*** participation   (RA33_C53 M3S1A + M6S2A)
also ***enable students to make*** more info   (RA20_21 M6S2B)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
to vb nns pos	**13**	**6**	**7.35108**
(to-verb, base form- noun, plural-possessive ending)			

M6S1

to check participants ' knowledge   (RA17_73 M6S1B)
M6S2
to discern readers ' comprehension   (RA18_C40 M6S2B)
****to draw learners*** ' attention to   (RA17_48M6S2A)
to enhance students ' retention   (RA2_15 M6S2A)
to enhance students ' vocabulary   (RA12_C35 M6S2A)
To enhance students ' perceptions   (RA14_29 M6S2B)
to enhance students ' intrinsic m   (RA14_C46 M6S2A)
to enhance students ' perceptions   (RA14_26 M6S2A)
to foster students ' perceptions   (RA14_C47 M6S2A)
to identify teachers ' reactions   (RA35_30 M6S2A)
****to increase students*** ' opportunities   (RA14_C47 M6S2A)
to maximize learners ' potential   (RA13_C21 M6S2B)
to promote students ' perceptions   (RA14_C46 M6S2A)

4-PoS tag Bundle in M6   f   Expected   Log Likelihood

	f	Expected	Log Likelihood
md vb vbn in	44	31	7.31872

(modal-verb, base form- verb- past participle- Preposition or subordinating conjunction)

M6S1

can be explored against the backdr op   (RA34_7 M6S1B)
***can be promoted in* Hong Kong   (RA36_C54 M6S1A)
***can be recruited from* the same   (RA34_C32 M6S1B)
can be thought of as such   (RA11_C71 M6S1A)
could be investigated under different   (RA8_C36 M6S1B)
could be presented with a transcri   (RA6_C56 M6S1B)
***could be tested in* a further   (RA18_C48 M6S1A)
***may be required for* the focus   (RA9_62 M6S1A)
***should be collected in* both dyadic   (RA3_C95 M6S1B)
should be conducted with a larger   (RA32_C31 M6S1B)
must be considered in studies on   (RA3_C74 M6S1A)
should be noted that this approach   (RA6_C62 M6S1B)
***should be taken into* consideration   (RA2_12M6S2A+M6S1A)

M6S2

can be argued that in the   (RA3_C70 M6S2B)
can be demonstrated by means of   (RA19_C49 M6S2A)
can be evaluated for its transfe   (RA4_49 M6S2A)
***can be filtered through* their be   (RA20_C37 M6S2A)
can be fostered through genuine in   (RA14_33 M6S2B)
***can be found between* oral and   (RA31_C36 M6S2A)
can be identified in relation to   (RA4_47 M6S2A)
can be introduced by teachers wit   (RA7_C56 M6S2A)
can be linked in the learner '   (RA4_46 M6S2A)
***can be linked in* the learner '   (RA4_48 M6S2A)
can be suggested that teachers in   (RA20_19 M6S2A)
can be suggested that teachers sh   (RA12_C35 M6S2A)
can be speculated that the teacher   (RA20_C37 M6S2A)
can be targeted in task-based   (RA33_16 M6S2A)
can be translated into simple advi   (RA19_C48 M6S2B)
***could be addressed by* engaging   (RA28_61 M6S2A)
***could be reinforced if* tasks are   (RA31_21 M6S2A)
could be used by the proofreader   (RA26_C56 M6S2A)
***may be addressed by* revising the   (RA35_31 M6S2A)
may be combined with other methods   (RA16_C58 M6S2B)
***may be translated into* pedagogical   (RA27_C24 M6S2B)
***might be achieved by* opening a   (RA36_35 M6S2A)
***might be achieved by* providing NE   (RA36_35 M6S2A)
***might be suggested that* applying   (RA19_C41 M6S2A)
***should be provided with* social eve   (RA10_79 M6S2A)
**s*hould be provided with* information   (RA35_28 M6S2A)
should be recognized as an integra   (RA20_C38M6S2A+M4S1)
***should be selected with* a view   (RA18_C36 M6S2A)
should be shared with all the   (RA33_16 M6S2A)

should be used as a template ,   (RA24_C63 M6S2A)
*will be recalled tha**t** our inform   (RA26_C58 M6S2A)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
md vb jj nns	12	5	6.58555

(modal-verb, base form- adjective-noun, plural)

M6S1
would provide valuable insights into   (RA1_30 M6S1A)
might yield different results for   (RA14_C41 M6S1B)
*would reveal different stances due   (RA23_C63 M6S1A)
M6S2
can include myriad emotions , frustrations   (RA29_42 M6S2A)
can make authentic texts more acc   (RA6_C51 M6S2A)
*can represent diverse cultures , no   (RA25_C57 M6S2A)
should invite guest speakers to th   (RA10_72 M6S2A)
*should provide multiple opportunities for   (RA14_31 M6S2A)
*should provide multiple opportunities for   (RA35_25 M6S2A)
*should use individual criteria to   (RA14_29 M6S2B)
*would be effective means.   (RA32_17 M6S2A)
*would permit such exchanges to tak   (RA26_12 M6S2B)

4-PoS tag Bundle in M6	f	Expected	Log Likelihood
nns md vb dt	17	9	6.38586

(noun, plural- modal-verb, base form- determiner)

M6S1
*factors may outweigh the influence   (RA31_24 M6S1A)
*measures would be an appropriate f   (RA11_C69 M6S1B)
*researchers should use a research   (RA12_C38 M6S1B)
*studies should compare the roles   (RA10_C104 M6S1A)
*studies should examine the relationsh   (RA10_C105M6S1A)
studies should examine the long t   (RA10_C105 M6S1A)
*studies will find a way to   (RA2_19   M6S1B)
tasks would be another interesting   (RA2_18 M6S1A)
M6S2
*authorities will wish the writer '   (RA26_C51 M6S2A)
*listeners can benefit the most fr   (RA6_C48 M6S2A)
*narratives may be a particularly b   (RA9_55 M6S2A)
*students can be an effective activ   (RA2_13 M6S2A)
teachers may lack both the knowled   (RA21_C56 M6S2A)
*teachers should create a classroom   (RA14_33 M6S2B)
*teachers should adopt these disciplinary   (RA32_23 M6S2A)
*universities would have an overview   (RA26_C45 M6S2A)
*users should develop an awareness   (RA32_24 M6S2B)

---

*Part of them was captured in other PoS bundles.

S.   Formulaic Move 6 bundles having modals

	M6S1 **5-TAG**	**M6S1** **4-TAG**	**M6S2** **5-TAG**	**M6S2** **4-TAG**
can	*can be done to promote* *can affect the usefulness of* *teachers can be recruited from* *stakeholders can be promoted* *in*	*training programs can help* *brief remarks can be* **different stakeholders can be* *following areas can help* *subsequent revisions can be* *such research can contribute* **can affect the usefulness* *engagement can deepen* *underst* *can be explored against* **can be promoted in* **can be recruited from* *teachers can be kept* *can be thought of*	*can be asked to consider* *can be devised to focus* *addition, instructors can utilize* *behavior, teachers can promote* *instruction, learners can benefit* *time, instructors can provide* *teachers can try to understand* *can increase the amount of* *can provide a means for* *decisions can be filtered throug* *event can stimulate the writer*	*L2 students can be* *corpus tools can help* *ones can be copied* *individual instructors can do* **instructional tasks can be* *L2 teachers can draw* *natural sciences can be* *prepared materials can provide* *skilled listeners can benefit* *specific task can be* *such work can be* *appropriate balance can be* *can elect to focus* **can try to understand* *can help the students* *can increase the amount* **can provide a means* *event can stimulate the* *topics can represent diverse* **can represent diverse cultures* *instructors can supplement whole* *instructors can do more* *instructors can utilize specific* **sciences can be problematic* *study can provide a* **decisions can be filtered*

could

could help the selection of
could encompass an
examination of
*inquiry could encompass an
examination
Moreover, further studies
could

present finding could be
present study could be
current study could be
vocabulary learning could
include
*could encompass an
examination
could help the selection
context could include teachers'
study could have implications
could be investigated under
could be presented with

could be instructed to contact
could be found to help
could become the basis for
team could provide a forum
exams could be reinforced if
goals could be addressed by
Moreover, social activities could
approach could include the
following

can be speculated that
can be argued that
can be demonstrated by
can be evaluated for
can be fostered through
can be identified in
can be introduced by
can be linked in
can be suggested that
can be suggested that
can be targeted in
can be translated into
can include myriad emotions
can make authentic texts
* repertoire goals could be
*discourse norms could be
explicit links could be
* social activities could be
such input could include
* could become the basis
* could include the following
team could provide NETs
team could provide a
* could provide a forum that
* input could include aspects
approaches could be feasible
* approach could include the
* activities could be introduced
* exams could be reinforced
links could be built
practices could be enhanced
students could be found

234

may	may be necessary to examine may be required to oust may be enhanced to increase researchers may need to examine may outweigh the influence of	point learners may become *factors may outweigh the instructional methods may be *may need to examine *may be necessary to input may require a *may be required for reference may be the	may be able to draw may be able to pass correctness, teachers may achieve teachers may resort to writing may include an identification of study may provide a way concerns may be addressed by	*could be addressed by could be used by * language teachers may resort heritage speakers may benefit appropriate narratives may be general suggestions may serve task-based courses may be conventionalized genre may provide * own disciplines may be * current study may include professional genre may be may need to be may include an identification may provide a way * may resort to writing products may obtain lower teachers may achieve better learners may have few * disciplines may be able * may be able to * study may include an * study may provide a teaching may affect the model may be a approach may demand a * concerns may be addressed may be combined with teachers may lack both
might	might be worthwhile to analyze might be hard to examine	*Future research might probe future research might use student observations might	environment , researchers might begin teachers might seek to increase	Brazilian researchers might write Brazilian authors might be elementary schools might be

researchers might choose to write
*researchers might begin to investigate
might probe this issue with
research might use an eye tracking
might be an avenue for
work might increase the learners'

provide
*further work might show
*work might increase the
ethnographic study might help
*might begin to investigate
might choose to write
*might be an avenue
*might be hard to
*might be worthwhile to
activities might be increased
might yield different results

* English teachers might seek
positive results might inspire
extra step might provide
useful strategy might be
critical consciousness might be
professional development might benefit
might help to convince
might influence a writer
* schools might be difficult
such problematization might be
*might be achieved by
*might be suggested that

should

should be administered to ask
should be explored to determine
research should explore the beliefs
Furthermore, future studies should
Therefore, future studies should
Therefore, vocabulary researchers should
data should be collected in
factors should be taken into

future studies should examine
*future studies should examine
*vocabulary researchers should use
future research should examine
*word factors should be
Further research should investigate
Further research should explore
*future studies should compare
studies should examine the
*should explore the beliefs
Such analyses should be
*data should be collected
*tasks should be explored

should be collected to identify
should be encouraged to use
should be encouraged to observe
should be permitted to work
competence, teachers should use
motivation , instructors should attempt
course, departments should provide
orientation, participants should be
reading, testers should be
instructors should attempt to incorporate
teachers should strive to provide
should consider the effectiveness of
should develop an awareness of

* topics should be selected
comprehension questions should be
* testers should be sensitive
student teachers should adopt
* should develop an awareness
*Learners should be
* should strive to provide
teachers should use individual
* should create a classroom
L2 educators should pay
key distinction should be
* should attempt to incorporate
should continue to encourage
* should consider the effectiveness
institution should arrange workshops
should be shared with

	variables should be kept	should discuss the relationship between	* students should be encouraged
	should be conducted with	institution should consider the effectiveness	voices should be heeded
	should be noted that	Learners should be provided with	should discuss the relationship
		participants should be provided with	should provide an impetus
		topics should be selected with	administration should find ways
		,we should provide multiple	staff should help students
			departments should provide multiple
			* teachers should be encouraged
			* data should be collected
			* proofreaders should be permitted
			*Learners should be provided
			* participants should be provided
			*should be provided with
			should be recognized as
			should be used as
			should invite guest speakers
			*should provide multiple opportunities
			*should use individual criteria
			*teachers should create a
			*teachers should adopt these
			*users should develop an
would	would be important to examine	collaborative task would provide	would be able to predict	would have an overview
	it would be interesting to	morphology measures would be	would be unable to do	extent universities would permit
	would be interesting to see	verbal interactions would reveal	materials would seem to provide	* prepared materials would seem
	would be interesting to learn	Such data would be	, it would be helpful for	genre analysis would be
	would be interesting to understand	Brazilian researchers would		reflective discourse would enable
	would be necessary to establish			would seem to be
				would seem to support
				would represent a learning

237

would be worthwhile to
proceed
it would be important to

help
Brazilian scholars would cast
L1 composition would have
cultural capital would enjoy
would continue to degrade
would need to be
would help to explore
would allow the students
*would be important to
*would be interesting to
would be interesting to
would be interesting to
would be interesting to
*would be necessary to
would be worthwhile to
*capital would enjoy a
idea would be a
participants would be
necessary
would provide valuable
insights
tasks would be another

research would enable teachers
* universities would permit such
guidelines would be beneficial
* would be able to
would be unable to
* would seem to provide

will	*will be important to quantify* *learners will need to be* *will be the subject of*	*Additional studies will help* *L2 learners will need* *Such research will allow* *empirical research will help* *such engagement will be* *will be the subject* *future studies will find* *work will inform efforts* **differences will be important*	*will foster a sense of* *connection will benefit all* *students*	**will benefit all students,* ** will foster a sense* *will wish the write*r's ** learning will foster a* *will be recalled tha*t
must		*must be considered in* *effects must be considered*		*such policies must be* *must continue to learn*

239