A CORPUS-BASED STUDY OF MALAYSIAN ESL STUDENTS' USE OF DISCOURSE CONNECTORS IN UPPER AND POST-SECONDARY ARGUMENTATIVE WRITING

By

SEYED ALI REZVANI KALAJAHI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirement for the Degree of Doctor of Philosophy

July 2014
Dedicated to

My parents, who taught me to live
My wife, who taught me to love
My supervisor, who taught me to learn
Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

A CORPUS-BASED STUDY OF MALAYSIAN ESL STUDENTS’ USE OF DISCOURSE CONNECTORS IN UPPER AND POST-SECONDARY ARGUMENTATIVE WRITING

By

SEYED ALI REZVANI KALAJAHI

July 2014

Chairman: Associate Professor Ain Nadzimah Abdullah, PhD
Faculty: Educational Studies

Discourse connectors (DCs) are one of the elements of cohesive devices that bring about cohesion to a piece of writing or speech. They are potentially useful means for writers, particularly in ESL and EFL writing pedagogic settings. DCs usefulness is two-pronged. First, they help and guide readers through the text, and then they are tools for writers to engage with their readers. It has been well-documented that appropriate and efficient use of DCs will create a coherent flow of the text. However, second/foreign language learners have some difficulties to use them efficiently and systematically in their writing. Literature review shows that the Malaysian ESL students are also suffering from improper and efficient use of DCs which leads them in failing to produce a cohesive text. Surprisingly, no single study was found in the context of Malaysia to investigate Malaysian ESL students’ understanding and use of DCs.

Hence, this study attempted to investigate and understand the nature and the use of the DCs in the Malaysian student writing compared with Native speakers writing. The study also was set to examine the correlation between the frequency use of the DCs and the quality of writing. The final goal of this research study was to find out to what extent Malaysian ESL students are committing errors while using DCs.

A corpus-based approach was adopted to meet the objectives of the study. To this end, an argumentative topic was assigned to the Form 4, Form 5 (upper-secondary) and the first year college students (post-secondary) and they were asked to write about the given topic in the classroom and submit their works to the instructors. They were required to write 250 words within 45 minutes. Upon compilation of the essays, the Malaysian Corpus of Students' Argumentative Writing (MCSAW) was built with ≃ 600,000 tokens. To compare and find out a vivid picture of Malaysian ESL students use of DCs with Native English Speakers, the Louvain Corpus of Native Essay Writing [LOCNESS] corpus was used.

Oxford Wordsmith Tools (5) was employed to extract data from corpus for analysis, by using frequency count and concordance functions. Aiming to identify what type of DCs is used by Malaysian ESL students, Discourse Connector List developed by Rezvani...
Kalajahi and Neufeld (2014) was used. To be able to examine the relationship between the quality of writing and the frequency of the use of the DCs, ESL composition profile offered by Jacobs et al. (1981) was utilized. Finally, a framework of identification of DCs error type was developed by the researcher to explore the errors that students commit while using DCs.

Findings of this study entail three phases. First, it was observed that Malaysian students tend to use DCs more frequently than native students. The overall frequency of the use of the DCs between Malaysian and native students was statistically significant at $p < .05$. However, the native students used more variety of DCs types than Malaysian students (398 vs. 328). It was also found that Malaysian students use DCs in some categories frequently and infrequently. Based on the findings in the native students writing (LOCNESS Corpus), the most frequent DCs in written English were offered. Second, there was a very weak negative but insignificant correlation between writing quality and the frequency of the use of the DCs in the writing of Malaysian ESL students. Finally, the qualitative analysis revealed that the erroneous use of DCs made by Malaysian ESL student writers mainly manifested in eight different categories. They had problems with the use of these devices which involved semantic, syntactic, stylistic, positional and mechanical errors. They also appeared to have tendency for unnecessary addition, omission, and redundant repetition of the DCs.

In conclusion, this study demonstrated that Malaysian ESL students’ use of DCs was still at an evolving level. It is vitally important that the accurate use of DCs in writing among Malaysian students be further highlighted in the classrooms through using concordance lines and adopting explicit instruction technique. Besides, material developers may take the outcome of the research into consideration and could find out possible ways to distribute and introduce DCs systematically across the educational levels.
Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah.

KAJIAN BERASASKAN KORPUS TENTANG PENGGUNAAN PENGHUBUNG WACANA DALAM KALANGAN PELAJAR ESL MENENGAH ATAS DAN PASCA MENENGAH MALAYSIA DALAM PENULISAN PERBINCANGAN BAGI

Oleh

SEYED ALI REZVANI KALAJAHII

Julai 2014

Pengerusi: Prof. Madya Ain Nadzimah Abdullah, PhD
Fakulti: Pengajian Pendidikan

Penghubung wacana - Discourse Connectors (selepas ini DC) merupakan salah satu daripada unsur-unsur peranti kohesif membawa kepada penulisan atau ucapan yang padu. Perkara ini sangat berguna bagi penulis, terutamanya dalam pedagogi penulisan ESL dan EFL. DC dianggap sebagai serampang dua mata. Pertama, mereka membantu dan membimbing pembaca dalam pembacaan teks, kemudian menjadi alat untuk penulis melibatkan diri dengan pembaca mereka. Hal ini telah didokumenkan dengan baik bahawa kesesuaian dan kecekapan penggunaan DC akan mewujudkan aliran koheren teks. Walau bagaimanapun, pelajar bahasa kedua/pelajar bahasa asing menghadapi kesukaran menggunakannya dengan cekap dan sistematis dalam penulisan mereka. Satu sorotan literatur menunjukkan bahawa pelajar ESL Malaysia juga mengalami penggunaan DC yang kurang cekap dan kurang berkesan yang kemudian mengakibatkan kegagalan untuk menghasilkan teks yang padu. Yang menghairankan, tiada pun satu kajian dijalankan dalam konteks Malaysia untuk mengkaji kefahaman dan penggunaan DC oleh pelajar ESL Malaysia.

Oleh itu, kajian ini bertujuan untuk mengkaji dan memahami sifat dan penggunaan DC bagi pelajar Malaysia berbanding dengan penutur asli. Kajian ini juga telah direka untuk mengkaji korelasi antara kekerapan penggunaan DC dan kualiti penulisan. Matlamat akhir kajian ini adalah untuk mengetahui sejauh mana pelajar ESL Malaysia melakukan kesilapan semasa menggunakan DC.

Pendekatan berasaskan korpus telah digunakan bagi mencapai objektif kajian. Untuk tujuan ini, satu topik penulisan perbincangan ditugaskan kepada pelajar tingkatan 4, Tingkatan 5, dan pelajar kolej tahun pertama. Pelajar-pelajar diminta untuk menulis tentang topik yang diberikan di dalam kelas dan menghantar karya mereka kepada pengajar. Mereka dikehendaki menulis sebanyak 250 patah perkataan dalam masa 45
minit. Setelah selesai penulisan itu disusuni, korpus penulisan perbincangan bagi pelajar-pelajar Malaysia (MCSAW) telah dibina dengan \( \approx 600,000 \) token. Untuk perbandingan dan memperolehi gambaran yang jelas antara pelajar-pelajar ESL Malaysia dengan penutur asli bahasa Inggeris dalam menggunakan DC, pendekatan Louvain Corpus of Native Essay Writing [LOCNESS] telah digunakan.


Hasil kajian ini melibatkan tiga fasa. Pertama, ia telah memperlihatkan bahawa pelajar Malaysia menggunakan DC lebih banyak berbanding dengan pelajar-pelajar penutur asli. Kekerapan keseluruhan penggunaan DC antara pelajar Malaysia dan pelajar penutur asli didapti mempunyai statistik yang ketara pada \( p < .05 \). Walau bagaimanapun, pelajar penutur asli menggunakan pelbagai jenis DC berbanding dengan pelajar Malaysia (398vs.328). Kajian ini juga mendapati bahawa terdapat pelajar-pelajar Malaysia yang menggunakan beberapa kategori DC secara berlebihan dan berkurangan. Berdasarkan penemuan kajian ini penulisan pelajar penutur asli (LOCNESS Corpus), yang paling kerap menggunakan DC telah ditawarkan. Kedua, terdapat korelasi yang menunjukkan hubungan yang negatif dan lemah tetapi tidak ketara antara kualiti penulisan dan kekerapan penggunaan DC dalam penulisan pelajar ESL Malaysia. Akhir sekali, analisis data kualitatif pula menunjukkan bahawa penggunaan DC yang salah telah dilakukan oleh penulis pelajar ESL Malaysia terutamanya dimanifestasikan dalam lapan kategori yang berbeza. Mereka mempunyai masalah tentang penggunaan alat-alat ini yang melibatkan semantik, sintaksis, gaya, kedudukan dan kesilapan mekanikal. Mereka juga mempunyai kecenderungan untuk penambahan yang tidak perlu, tidak menggunakan, dan pengulangan penggunaan DC.

Kesimpulannya, kajian ini menunjukkan bahawa penggunaan DC oleh pelajar ESL Malaysia masih pada tahap perkembangan. Ia amat penting untuk dititikbertkan penggunaan DC yang tepat dalam bilik darjah melalui penggunaan talian konordans dan menerima pakai teknik arahan-arahan terperinci. Selain itu, pemaju bahan pengajaran boleh mempertimbangkan hasil kajian ini dan mengenalpasti pelbagai pelaksanaan untuk mengedarkan dan memperkenalkan DC secara sistematik di seluruh peringkat pendidikan.
ACKNOWLEDGEMENTS

I would like to express my gratitude to a number of people for their help, support, and cooperation in conducting this dissertation.

First of all, I would like to thank my supervisor, Associate Professor Dr. Ain Nadzimah Abdullah who took her precious time and rendered me valuable advice and constructive comments in writing this thesis. She was a great help in shaping my study. A profound gratitude is also owed to the feedback of the committee members, Professor Dr. Jayakaran Mukundan and Associate Professor Dr. Arshad Abd. Samad for their valuable recommendations, worthwhile suggestions and encouragement.

Special thanks go to my wife and my family for their constant support and encouragement all through my life when I needed it most.

My thanks are also expressed to Dr. Vahed Zarifi and Dr. Vahid Nimechisalem, Bakhtiar Naghdipour, and Steve Neufeld for their feedback and rating the students’ essays. I am also grateful to the teachers and students who patiently helped me in building the MCSAW corpus.

Last but not least, I am truly grateful to Allah for giving me patience, strength, and wisdom to complete this thesis.
I certify that a Thesis Examination Committee has met on 18 July 2014 to conduct the final examination of Seyed Ali Rezvani Kalajahi on his thesis entitled "A Corpus-Based Study of Malaysian ESL Students' use of Discourse Connectors in Upper and Post-Secondary Argumentative Writing" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

**Ismi Arif bin Ismail, PhD**  
Associate Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Chairman)

**Chan Swee Heng, PhD**  
Professor  
Faculty of Modern Language and Communication  
Universiti Putra Malaysia  
(Internal Examiner)

**Nooreen binti Noordin, PhD**  
Senior Lecturer  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Internal Examiner)

**Randi Reppen, PhD**  
Professor  
Northern Arizona University  
United States  
(External Examiner)

NORITAH OMAR, PhD  
Associate Professor and Deputy Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date: 18 August 2014
This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee are as follows:

**Ain Nadzimah Abdullah, PhD**  
Associate Professor  
Faculty of Modern Languages and Communication  
Universiti Putra Malaysia  
(Chairman)

**Arshad Abd. Samad, PhD**  
Associate Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
<Member>

**Jayakaran Mukundan, PhD**  
Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
<Member>

**BUJANG BIN KIMHUAT, PhD**  
Professor and Dean  
School of Graduate Studies  
Universiti Putra Malaysia  

Date:
DECLARATION

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ______________________  Date: ________________

Name and Matric No: Seyed Ali Rezvani Kalajahi GS29486
Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: [Signature]
Name of Chairman of Supervisory Committee: [Name of Chairman]
[Name of Committee: Associate Professor Dr. Ain Nadzimah Abdullah]

Signature: [Signature]
Name of Member of Supervisory Committee: [Name of Member]
[Name of Committee: Assoc. Prof. Dr. Arshad Abd. Samad]

Signature: [Signature]
Name of Member of Supervisory Committee: [Name of Member]
[Name of Committee: Prof. Dr. Jayakaran Mukundan]
TABLE OF CONTENTS

ABSTRACT I
ABSTRAK III
ACKNOWLEDGEMENTS V
APPROVAL VI
DECLARATION VIII
LIST OF TABLES XIII
LIST OF FIGURES XIV

CHAPTER

1 INTRODUCTION 1
1.1 Background of the Study 1
1.2 Discourse Connectors 3
1.3 Corpus and its Connection to Language Teaching 5
1.4 Nature of the Present Study 6
1.5 Rationale of the Present Study 7
1.6 Statement of the Problem 8
1.7 Objectives 11
1.8 Research Questions 11
1.9 Significance of the Study 12
1.10 Limitations 12
1.11 Definition of Terms 14

2 LITERATURE REVIEW 17
2.1 Theoretical Underpinnings 17
  2.1.1 Theories of Writing 17
    2.1.1.1 Social Constructivism Theory 17
    2.1.1.2 Social Cognitive Theory 18
  2.1.2 Two Approaches to DCs 21
    2.1.2.1 Coherence-based Approach 21
    2.1.2.2 Relevance-theoretic Account Approach 23
      2.1.2.2.1 Relevance Theory and DCs 24
  2.1.3 Conclusion of Two Approaches 26
  2.1.4 Theoretical Framework of the Study 27
2.2 Term, Definition and Taxonomy of DCs in the Literature 28
  2.2.1 What are DCs? 34
  2.2.2 DCs List 36
    2.2.2.1 Description of DCs List 39
2.3 Text Coherence and DCs 41
  2.3.1 The Importance of DCs in Written Discourse 41
  2.3.2 The Sources of Difficulty in Learning DCs 43
2.4 Error Analysis 47
  2.4.1 Relevance of Error Analysis in English Language Teaching 49
  2.4.2 Model for Error Analysis 50
    2.4.2.1 Sources of Errors 50
    2.4.3 Advantages of Error Analysis 51
2.5 Corpus Linguistics 52
  2.5.1 What is a Corpus? 52
2.5.2 Advantages of Using Computers 52
2.5.3 Nature of Corpus Linguistics 53
2.5.4 Corpus Processing 53
   2.5.4.1 Input of Information to a Corpus 53
   2.5.4.2 Output of Information from a Corpus 54
2.5.5 Approaches to Corpus Linguistics 54
   2.5.5.1 Corpus-based Approach 54
   2.5.5.2 Corpus-driven Approach 55
2.5.6 Characteristics of a Modern Corpus 55
2.5.7 Types of Corpora 56
   2.5.7.1 Learner Corpus 56
      2.5.7.1.1 Need for a Learner Corpus 58
         2.5.7.1.1.1 Corpus Work in Malaysia 60
   2.5.7.2 Reference Corpora 61
2.6 Previous Studies on DCs in Different Settings and Genres 62
2.7. Conclusion 76

3 METHODOLOGY 77
3.1 Corpus Design 77
   3.1.1 Purpose of Building the MCSAW Corpus 77
   3.1.2 Register of Text 78
   3.1.3 Task Setting 78
   3.1.4 Genre and Topic Choice 78
   3.1.5 Essay Length 79
   3.1.6 Corpus Size and Subjects 80
   3.1.7 Corpus Sampling: Representativeness and Generalization 81
   3.1.8 Corpus Collection Procedure 82
3.2 Reference Corpus 82
3.3 Research Design 82
3.4 Identification of DCs 84
3.5 Method of Data Analysis 85
   3.5.1 Instrumentation 86
      3.5.1.1 Wordsmith 86
      3.5.1.2 ESL Composition Profile 87
   3.5.2 Data Analysis 88
      3.6.1 Quantitative Phase 88
         3.6.1.1 Frequency Count and Proportion Analysis 88
         3.6.1.2 Discrepancy Analysis 88
         3.6.1.3 Correlational Analysis 89
      3.6.2 Qualitative Phase 90
         3.6.2.1 Identification and Classification of Error Use of DCs 90
3.7 Research Framework 92
3.8 Conclusion 94

4 RESULTS AND DISCUSSION 95
4.1 Frequency of Occurrence 95
   4.1.1 Overall Frequencies 95
   4.1.2 Individual DCs 100
   4.1.3 Position of DCs 105
4.1.4 Frequently and Infrequently Used DCs 107
4.2 Frequency of Semantic Types of DCs 110
4.3 The Most Frequent DCs of Written English 118
4.4 Relationship between Quality of Writing and Frequency of DCs 121
  4.4.1 Summary and Discussion 123
4.5 Error Analysis of DCs in Malaysian Students’ Writing 124
  4.5.1 Error Types 125
    4.5.1.1 Semantic Misuse of DCs 125
    4.5.1.2 Syntactic Misuse of DCs 126
    4.5.1.3 Stylistic Misuse of DCs 130
    4.5.1.4 Positional Misuse of DCs 135
    4.5.1.5 Unnecessary Addition of DCs 137
    4.5.1.6 Omission of DCs 137
    4.5.1.7 Redundancy in the use of DC 140
    4.5.1.8 Mechanical Misuse of DCs 142
  4.5.2 Summary and Discussion 143
4.6 Concluding Remarks 147

5 CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS 148
  5.1 Summary of Findings, Discussions and Conclusions 148
  5.2 Implications of the Study 151
  5.3 Recommendations for Future studies 154
  5.4 Contributions of the Study 155
    5.4.1 Assigning DC to Primary and Secondary Levels of Education 155
  5.5 Concluding Remarks 156

REFERENCES 157
APPENDICES 177
  Appendix A : Discourse Connectors List 177
  Appendix B: LOCNESS Description 201
  Appendix C: ESL Composition Profile 208
  Appendix D: List of the Sorted and the Ranked DCs 210
  Appendix E: List of the Common DCs between MCSAW and LOCNESS with an Approximate Frequency of Occurrence Per Million 234
  Appendix F: Pie Chart for Distribution of the DCs Categories within Sub-corpus 236
  Appendix G: Distribution of the DCs Sub-categories within Sub-corpus 239
  Appendix H: Two Raters Scores and DCs Frequencies in the Selected Writing 240
  Appendix I: The Discourse Connector List mapped into the primary and secondary levels in Malaysia (Standard 1-6 & Form 1-5) 243

BIODATA OF STUDENT 251
LIST OF PUBLICATIONS 252