



Ain Shams University

Faculty of Women for Arts, Science and Education

Department of Curricula & Instruction

***A Corpus-based Program for Developing the Phraseological
Competence of Sinai University Media Students***

A Dissertation

Submitted for the PhD Degree in Education

(Curricula and Methods of Teaching English as a Foreign Language)

by

Amal Salama Nasrallah Nasr

Under the Supervision of

Dr. Aida Abdel Maksoud Zaher

Professor of Curricula and Methods of
Teaching English as a Foreign Language
Faculty of Women
Ain Shams University

Dr. Ahmed Anwar Thabet

Associate Professor of Linguistics
Faculty of Education
Mansoura University

Dr. Manal Muhammad Abdel Aziz

Associate Professor of Curricula & Methods of
Teaching English as a Foreign Language
Faculty of Women
Ain Shams University

2015

Abstract

Nasrallah, Amal Salama. A Corpus-Based Program for Developing the Phraseological Competence of Sinai University Media Students. Unpublished PhD Dissertation. Women College, Ain Shams University, 2015.

The current study aims at investigating the effectiveness of a corpus-based program for developing the phraseological competence of Sinai University Media Students (SUMS). The study adopts the one-experimental group design. Consequently a group of 40 media students were selected and received the program. This corpus-based program is based mainly on students' phraseological needs that were defined through a corpus-based analysis based on Concgram 0.1 software. It tests if there are statistically significant differences between the mean scores of the treatment group's students on the pre and post administration of the phraseological competence test of learning phraseological units of nouns, verbs and prepositions in favor of the post administration in addition, if there are statistically significant differences between the mean scores of the treatment group's students on the pre and post administration of the writing test of using phraseological units of nouns, verbs and prepositions in favor of the post administration. The study makes use of corpus-based tools of data collection, corpus-based tools of data analysis, a pre-post test to measure SUMS progress in learning phraseological units of nouns, verbs and prepositions, and a pre-post writing test to measure SUMS use of phraseological units of nouns, verbs and prepositions. The corpus-based program was taught over a period of three months. By the end of the treatment, the study tests were administered to the treatment group. The post-administration revealed the effectiveness of the corpus-based program in developing the phraseological competence of the treatment group. The study results supported the study hypotheses, and the suggested program. As a result and based on such conclusions, a number of recommendations and suggestions for further research were presented.

Key words: Corpus linguistics – Learner corpus – Specialized corpus - Phraseology – Concgramming - ESP

Acknowledgments

I wish I could find the suitable words to express my deep thanks to Professor Aida Zaher, my supervisor, whose great guidance, help, tolerance and human treatment throughout the different stages of the research helped me accomplish it. I specially thank her for her constructive comments that always put the researcher on the right way. The thanks and gratitude are extended to Professor Manal Abdel Aziz who never hesitated to help me during the different stages of this work, giving from her precious time to support and aid me.

I would like to express my special appreciation and thanks to Professor Ahmed Thabet who has been a tremendous mentor for me. I would like to thank him for his care and support which allow me to grow as a researcher. His advice on the research as well as on my career has been priceless. I owe a great debt of gratitude to his gracious hospitality and tireless pursuit of excellence in the research that without his supervision and constant help this dissertation would not have been possible.

Thanks and appreciation are extended to the members of the committee; Dr. Afaf Abdel Hameed and Dr. Amira Khater for their precious advice and constructive comments. Furthermore, I would like to thank Dr. Badran Abdel-Hammed, Dr. Aly Qora, Dr. Iman El Bishbishi, Dr. Shimaa Torky and Dr. Hoda Abu Hashim; the jury members who validated the study's tools giving generously their time and expertise to better my work. In addition, I would like to gratefully acknowledge Dr. Saleh Muhammad Saleh who helped me in the statistical aspects of this work.

Furthermore, I would also like to acknowledge with much appreciation and gratitude the crucial role of Dr. Wahid Tohamy who never hesitated to help and guide me throughout the different stages of the research. Also, special thanks and appreciation goes to Dr. Amin El-Banna for helping me get the corpus analysis package from USA.

In addition, I would like to thank the Faculty of Media at Sinai University for the facilities they provided me to accomplish this work. Also, I thank my colleagues in the English Department at Sinai University for their love, help and support they gave me throughout the research.

Also, I would like to express my deepest gratitude to Mrs. Fathia Thabet for her constant love and generosity she offered me in the difficult stages of this research.

Finally yet importantly, I would like to thank my dear mother, my dear father and my dear sisters for their patience and love. If this research was blessed, their prayers would be the reason for that.

Table of Contents

Abstract.....	I
Acknowledgments	II
Table of Contents.....	III
List of Figures	Ix
List of Tables	X
List of Abbreviations	XIII

Chapter One: Introduction and Problem

1.1 Introduction.....	2
1.2 Context of the Problem	14
1.3 Statement of the Problem.....	17
1.4 Purpose of Study.....	18
1.5 Hypotheses of study	18
1.6 Variables of Study.....	19
1.7 Delimitations of Study.....	20
1.8 Significance of Study	21
1.9 Definition of terms	22

Chapter Two: Theoretical Background

2.1 Corpus Linguistics	30
2.1.1 Origins and Evolution of Corpus Linguistics	31
2.1.2 The Term 'Corpus' in Corpus Linguistics	33

2.1.3 Kinds of data corpora present	35
2.1.4 Types of corpora	36
2.1.5 Corpus Construction	38
2.1.5.1 Designing a Specialized Corpus	40
2.1.5.2 Designing a Learner Corpus	42
2.1.6 Corpus Analysis	43
2.1.6.1 Basics of Corpus Analysis	43
2.1.6.2 Procedures of Corpus Analysis	44
2.1.6.3 Corpus Analysis Tools	46
2.1.6.4 Statistics in Corpus Linguistics	47
2.1.7 Corpus Linguistics: An Empirical Methodology	50
2.1.8 Corpus-based Multidisciplinary Research	51
2.1.9 Corpus Linguistics in English Language Teaching	52
2.1.9.1 Direct Applications of Corpora in Language Teaching (DDL)	55
2.1.9.2 Indirect applications of corpora in language teaching...	64
2.2 Phraseology	69
2.2.1 Phraseology and phraseological units	71
2.2.2 Typology of phraseological units	74
2.2.3 Corpus-based analysis of phraseological units	
2.2.4 Phraseology in foreign language learning and teaching	84
2.2.5 Phraseology in specialized registers	87
2.3 English for Specific Purposes (ESP)	89

2.3.1 Definition of ESP	91
2.3.2 Areas of ESP research	92
2.3.3 Corpus analysis and ESP	94
2.4 Needs analysis	98
2.4.1 Needs defined	98
2.4.2 Needs analysis defined	99
2.4.3 Data collection for needs analysis	100
2.4.4 Needs analysis in Corpus linguistics	100
2.5 Concluding remarks.....	102

Chapter Three: Review of Related Studies

3.1 ESP learners' phraseological needs	105
3.1.1 Non-native Learner corpora in ESP Needs Analysis.....	106
3.1.2 Native Specialized Corpora in ESP Learners' Needs Analysis.	110
3.2 Corpora and ESP Phraseology	114
3.3Corpora and Data-driven Learning in ESP	119
3.4 Concluding remarks	127

Chapter Four: Method and Procedures

4.1 Design of the study.....	131
4.2 Participants	131
4.3 Tools of the study.....	132
4.3.1 Corpus-based tools for data collection	132
4.3.2 Tools of data analysis.....	137
4.3.3 The pre-post phraseological competence test.....	151

4.4 The writing-test.....	159
4.4.1 The test objective.....	159
4.4.2 Description of the test.....	159
4.4.3 Test validity.....	160
4.4.4 Piloting the test.....	160
4.4.5 Test Reliability.....	161
4.4.6 Scoring the test.....	161
4.4.7 Pre-post test administration.....	162
4.5 The Corpus-based program.....	162
4.5.1 Assumptions of the program.....	162
4.5.2 Procedures of designing the Corpus-based Program.....	165
4.5.3 Components of the corpus-based program in its final version..	168

Chapter Five: Data Analysis and Statistical Results

5.1 Corpus-based analysis of SUMS-C	180
5.1.1 SUMS-C tagging.....	180
5.1.2 Phraseological analysis of SUMS-C.....	182
5.1.2.1 Nouns phraseological profile.....	183
5.1.2.2 Verbs phraseological profile.....	200
5.1.2.3 Prepositions phraseological profile.....	214
5.2 SPSS Statistical analysis of the study hypotheses.....	230
5.3 Concluding Remarks	239

Chapter Six: Discussion of the Results, Conclusions and Recommendations

6.1 Discussion of the study results.....	241
6.1.1 Learner corpus as an innovative tool for analyzing ESP students' needs.....	241
6.1.2 Specialized native corpus as a pedagogic tool in learning phraseology.....	242
6.1.3 Findings of corpus-based phraseological analysis as empirical bases for ESP programs.....	242
6.1.4 Data-driven learning as an effective learning approach in ESP phraseology.....	243
6.1.5 Statistically significant results of the hypotheses.....	243
6.2 Conclusions.....	245
6.3 Recommendations.....	246
6.4 Suggestions for further research.....	247
References	249
 Appendices	
Appendix 1: English placement test's results of Media students.....	276
Appendix 2: The pilot Study.....	278
Appendix 3: Samples of SUMS writing.....	282

Appendix 4: Names of the jury who validated the phraseological competence test and the phraseological competence test in its final form....	287
Appendix 5: Names of the jury who validated the writing test and the phraseological competence test in its final form.....	296
Appendix 6: Names of the writing test raters.....	301
Appendix 7: Names of jury who validated corpus-based program.....	303
Appendix 8: The corpus-based program	305
 Arabic summary	 405

The attached CD

- 1. Sinai University Media Students Corpus (SUMS-C)**
- 2. Media Corpus of Native Speakers (MCNS)**
- 3. “Media” 2-word concgrams**
- 4. “news” 2-word concgrams**
- 5. “advertising” 2-word concgrams**
- 6. Information” 2-word concgrams**
- 7. “have” 2-word concgrams**
- 8. “get” 2-word concgrams**
- 9. “go” 2-word concgrams**
- 10. “Get” 2-word concgrams**
- 11. “in” 2-word concgrams**
- 12. “at” 2-word concgrams**
- 13. “on” 2-word concgrams**
- 14. “for” 2-word concgrams**

List of Figures

Figure No.	Title	Page
Figure (2.1)	The use of corpora in language learning and language teaching	53
Figure (2.2)	Cowie's (2001) classification of word combinations.....	74
Figure (2.3)	Mel'cuk's (1998) typology.....	75
Figure (2.4)	Burger's (1998) typology.....	75
Figure (2.5)	Distributional categories.....	77
Figure (2.6)	Areas of ESP (Jordan, 1997:3)	92
Figure (2.7)	ESP key types and sub-types Dudley Evans & Johns (1998: 6)...	92
Figure (2.8)	Needs analysis and corpus-based research in English language teaching.....	100
Figure (4.1)	The interface of Concgram 1.0	139
Figure (4.2)	An example of 2-word concgram	140
Figure (4.3)	Loading LC-SUMS	141
Figure (4.4)	Generating unique words.....	141
Figure (4.5)	Unique words of SUMS-C sorted by frequency order	142
Figure (4.6)	Defining SUMS phraseological profile procedures	143
Figure (4.7)	Automatic search of the 2-word	144
Figure (4.8)	Results shown in concgrams List Box	144
Figure (4.9)	T-scores of 'media' 2-word aboutgrams	146
Figure (4.10)	Generating 'media/mass' configuration	149

Figure (4.11)	Positional configuration of 'media/mass'	149
Figure (4.12)	Constituent configuration of 'media/mass'	150
Figure (4.13)	Summary configuration of 'media/mass'	150
Figure (4.14)	Media phraseological profile in SUMS writing	151
Figure (4.15)	SUMS phraseological needs.....	162
Figure (5.1)	Concordance lines for 'social/media' extracted from SUMS-C....	185
Figure (5.2)	Concordance lines for media/mass' extracted from SUMS-C	185
Figure (5.3)	Configuration list for 'media' 2-word Congrams	186
Figure (5.4)	Concordance lines for 'news/coverage' extracted from SUMS-C	189
Figure (5.5)	Concordance lines for 'news/channels' extracted from SUMS-C	190
Figure (5.6)	Configuration list for 'news' 2-word Congrams	190
Figure (5.7)	Concordance lines for 'information/get' extracted from SUMS-C	193
Figure (5.8)	Concordance lines for 'information/public' extracted from SUMS-C	194
Figure (5.9)	Configuration list for 'information' 2-word Congrams	195
Figure (5.10)	Concordance lines for 'advertising/agency' extracted from SUMS-C.....	198
Figure (5.11)	Concordance lines for advertising/products' extracted from SUMS-C	198
Figure (5.12)	Configuration list for 'advertising' 2-word Congrams	199
Figure (5.13)	Concordance lines for 'have/to' extracted from SUMS-C	202
Figure (5.14)	Concordance lines for 'have/part' extracted from SUMS-C	203

Figure (5.15)	Configuration list for 'have' 2-word Congrams	204
Figure (5.16)	Concordance lines for 'get/to' extracted from SUMS-C	206
Figure (5.17)	Concordance lines for 'get/news' extracted from SUMS-C	207
Figure (5.18)	Configuration list for 'get' 2-word Congrams	207
Figure (5.19)	Concordance lines for 'go/to' extracted from SUMS-C	210
Figure (5.20)	Concordance lines for 'go/in' extracted from SUMS-C.....	210
Figure (5.21)	Configuration list for 'go' 2-word Congrams	211
Figure (5.22)	Concordance lines for 'take/in' extracted from SUMS-C	213
Figure (5.23)	Configuration list for 'take' 2-word Congrams	213
Figure (5.24)	Concordance lines for 'interested/in' extracted from SUMS-C ...	216
Figure (5.25)	Concordance lines for 'in/between' extracted from SUMS-C	217
Figure (5.26)	Configuration list for 'in' 2-word Congrams	217
Figure (5.27)	Concordance lines for 'on/media' extracted from SUMS-C	220
Figure (5.28)	Concordance lines for 'on/news' extracted from SUMS-C	221
Figure (5.29)	Configuration list for 'in' 2-word Congrams	222
Figure (5.30)	Concordance lines for 'for/example' extracted from SUMS-C	225
Figure (5.31)	Concordance lines for 'for/fun' extracted from SUMS-C.....	225
Figure (5.32)	Configuration list for 'for' 2-word Congrams	226
Figure (5.33)	Concordance lines for 'at/last' extracted from SUMS-C	228
Figure (5.34)	Configuration list for 'at' 2-word Congrams	229
Figure (5.35)	The treatment subjects' mean scores on the phraseological competence pre-test and post- test	232
Figure (5.36)	The experimental group's mean scores on the pre-test and post-	233

	test of nouns phraseological units learning	
Figure (5.37)	The experimental group's mean scores on the pre-test and post- test of verbs phraseological units learning	234
Figure (5.38)	The treatment group's mean scores on the pre-test and post- test in written use of phraseological units	236
Figure (5.39)	The treatment group's mean scores on the pre-test and post- test in written use of nouns phraseological units	237
Figure (5.40)	The treatment group's mean scores on the pre-test and post- test in written use of verbs phraseological units	238
Figure (5.41)	The treatment group's mean scores on the pre-test and post- test in written use of prepositions phraseological units.....	239

List of Tables

Table No.	Title	Page
Table (2.1)	The difference between a ‘text’ and a ‘corpus’	35
Table (2.2)	Criteria of building learner corpora	43
Table (4.1)	Criteria of compiling SUMS-C.....	133
Table (4.2)	Criteria of compiling MCNS.....	135
Table (4.3)	Sources of MNSC	136
Table (4.4)	Parts of speech abbreviations in POS tagger.....	137
Table (4.5)	Top 4 most frequent nouns in LC-SUMS	142
Table (4.6)	Top 4 most frequent verbs in LC-SUMS	142
Table (4.7)	Top 4 most frequent prepositions in LC-SUMS ‘media’ 2- word concgrams of “media”	142
Table (4.8)	Test items that measure competence of each type of the phraseological units.....	154
Table (4.9)	Summary of the relation coefficients among individual raters.....	161
Table (4.10)	Criteria of scoring the writing test	161
Table (4.11)	Units the corpus-based program	171
Table (5.1)	Results of SUMS-C tagging	180
Table (5.2)	Frequency of LC-SUMS POS	182
Table (5.3)	‘media’ 2-word concgrams	183
Table (5.4)	‘media’ 2-word aboutgrams	183

Table (5.5)	T-scores of ‘media’ 2-word aboutgrams	184
Table (5.6)	"news" 2-word concgrams	187
Table (5.7)	‘news’ 2-word aboutgrams	187
Table (5.8)	T-scores of ‘news’ 2-word aboutgrams	188
Table (5.9)	‘information’ 2-word concgrams calculations	192
Table (5.10)	‘information’ 2-word aboutgrams	192
Table (5.11)	T-scores of ‘information 2-word aboutgrams	193
Table (5.12)	advertising 2-word concgrams calculations	196
Table (5.13)	‘advertising’ 2-word aboutgrams	196
Table (5.14)	T-scores of ‘advertising’ 2-word aboutgrams	197
Table (5.15)	‘have’ 2-word concgrams calculations	200
Table (5.16)	‘have’ 2-words aboutgrams	201
Table (5.17)	T-scores of ‘have’ 2-word aboutgrams.....	201
Table (5.18)	‘get’ 2-word concgrams calculations.....	204
Table (5.19)	‘get’ 2-words aboutgrams.....	205
Table (5.20)	t-scores of ‘get’ 2-words aboutgrams	205
Table (5.21)	‘go’ 2-word concgrams calculations	208
Table (5.22)	‘go’ 2-words’ aboutgrams	209
Table (5.23)	T-scores of ‘go’ 2-word aboutgrams	209
Table (5.24)	‘take’ 2-words concgrams	212
Table (5.25)	‘take’ 2-word aboutgrams	213
Table (5.26)	T-score of ‘take/in aboutgram	213
Table (5.27)	‘in’ 2-words concgrams	214