STATEMENT OF SOURCES

I declare that the work presented in the thesis is to the best of my knowledge and belief, original, except as acknowledged in the text, and that the material has not been submitted either in whole or in part for a degree at this or any other university.

Christina Gitsaki
ABSTRACT

This study examines the development of collocational knowledge in learners of ESL. A number of previous studies have underscored the importance of collocations for L2 acquisition, and the problems that learners face with learning and using collocations. However, there have been few attempts to systematically study how the development of collocational knowledge relates to the overall development of language proficiency with a particular intention in identifying possible stages in the development of L2 collocational knowledge. This study adopts a structure-based framework for the study of collocations based on previous studies (Zhang 1993; Biscup 1992) and the BBI Combinatory Dictionary of English, and attempts to describe how collocational knowledge develops across different language proficiency levels with respect to 37 collocation types. Data were collected from 275 Greek learners of ESL at three proficiency levels (post-beginners, intermediate, and post-intermediate) using three tasks: essay writing, translation test, and blank-filling. The essay writing provided evidence of accurate free production of collocations, while the translation and blank-filling tests measured accuracy in the subjects’ knowledge of collocations in cued production tasks. The data were examined with respect to the between- and within-group differences in accuracy on all three dependent measures. Statistical measures were employed to determine the significance of the observed between-group and within-group differences, and implicational scaling analyses was used to reveal accuracy orders in the acquisition of collocations. Results show that there are patterns of
development of collocational knowledge across and within the different proficiency groups for both the free and the cued production data. Collocational knowledge was shown to increase steadily as the level of proficiency increased, and there were group-specific accuracy orders showing that grammatical collocations are easier to acquire than lexical collocations. The development of collocational knowledge was found to be influenced by the syntactic complexity of the collocation types, and also by exposure and maturation.

Finally, three stages for the development of collocational knowledge are proposed. In the first stage learners acquire collocations as unanalysed lexical items, and hence the learners are more accurate with lexical collocations than complex grammatical ones. At the second stage the learners' grammatical knowledge develops enabling learners to use complex grammatical collocations with greater accuracy than in stage one. At the third stage of collocational development, learners are able to use both grammatical and lexical collocations with greater accuracy than in the other two stages, and they are on their way to a more advanced level of collocational knowledge.

Pedagogical implications and directions for future research are provided in light of the research findings.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATEMENT OF SOURCES</td>
<td>ii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xvii</td>
</tr>
<tr>
<td>ABREVIATIONS</td>
<td>xviii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>x</td>
</tr>
<tr>
<td><strong>CHAPTER 1: THE STUDY OF COLLOCATIONS</strong></td>
<td></td>
</tr>
<tr>
<td>1.0. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1. The Importance of Collocations in L2 Learning</td>
<td>4</td>
</tr>
<tr>
<td>1.2. Collocations in L2 Acquisition Research</td>
<td>13</td>
</tr>
<tr>
<td>1.3. Approaches to the Study of Collocations</td>
<td>22</td>
</tr>
<tr>
<td>1.3.1. The Lexical Composition Approach</td>
<td>22</td>
</tr>
<tr>
<td>1.3.2. The Semantic Approach</td>
<td>28</td>
</tr>
<tr>
<td>1.3.3. The Structural Approach</td>
<td>36</td>
</tr>
<tr>
<td>1.3.4. Summary of the Three Approach</td>
<td>46</td>
</tr>
</tbody>
</table>
CHAPTER 2: SECOND LANGUAGE ACQUISITION AND THE DEVELOPMENT OF COLLOCATIONAL KNOWLEDGE

2.0. Introduction................................................................................................55
2.1. Morphology.................................................................................................57
2.2. Syntax............................................................................................................61
2.3. Phonology....................................................................................................66
2.4. Vocabulary...................................................................................................67
   2.4.1. Vocabulary as a Language Sub-Skill...........................................68
   2.4.2. Vocabulary as a Language Skill...................................................70
   2.4.3. The Acquisition of Lexical Phrases.............................................75
   2.4.4. The Acquisition of Collocations..................................................78
   2.4.5. Summary..........................................................................................83
2.5. The Aims of the Present Study.................................................................84

CHAPTER 3: METHODOLOGY FOR THE PRESENT STUDY

3.0. Introduction................................................................................................86
3.1. Analysis of the Teaching Materials............................................................86
   3.1.1. Curriculum Objectives...................................................................87
   3.1.2. Syllabus and Methodology............................................................87
3.1.3. Activities and Tasks.................................................................88
3.1.4. The Use of L1........................................................................89
3.1.5. The Vocabulary.....................................................................90
3.1.6. Descriptive Statistics for the TWE Series............................90

3.2. Subjects.....................................................................................95

3.3. Materials....................................................................................96

3.3.1. Questionnaire......................................................................96
3.3.2. Composition.........................................................................97
3.3.3 Translation............................................................................97
3.3.4. Blank-Filling.......................................................................98

3.4 Data Collection Procedures....................................................100

3.4.1. Test Administration.............................................................100
3.4.2. Debriefing............................................................................102

3.5. Coding Procedures.................................................................102

3.5.1. Free Composition...............................................................103
   i) Holistic Rating......................................................................103
   ii) Target-Like Use of Articles...............................................104
   iii) Lexical Density.................................................................106
   iv) Terminal-Units....................................................................107
   v) Length of T-Units...............................................................108
   vi) Error-Free T-Units............................................................108
   vii) S-Nodes per T-Unit..........................................................109

3.5.2. Use of Collocations in the Essays.................................110

3.5.3. Translation.........................................................................111
3.5.4 Blank-Filling..........................................................111

3.6. Analyses..................................................................112

3.6.1. Language Proficiency Measures.........................112

3.6.2. Analyses for the Hypotheses..............................113

i) Analysis for Hypothesis 1......................................113

ii) Analysis for Hypothesis 2..................................114

CHAPTER 4: ANALYSES AND RESULTS

4.0. Introduction..........................................................116

4.1. Language Proficiency Results..............................118

4.1.1. Descriptive Statistics..........................................118

4.1.1.1. Descriptive Statistics for Group 1.................118

4.1.1.2. Descriptive Statistics for Group 2.................119

4.1.1.3. Descriptive Statistics for Group 3.................119

4.1.2. Results of the MANOVA.................................120

4.1.3. Holistic Rating................................................121

4.1.4. Target-Like Use of Articles............................121

4.1.5. Lexical Density...............................................122

4.1.6. Length of T-Units...........................................123

4.1.7. Error-Free T-Units.........................................125

4.1.8. S-Nodes per T-Unit...........................................126

4.1.9. Summary of the Results for the Language Proficiency Analyses..................................................127

4.2. Results of the Main Analyses.................................130
4.2.1. Hypothesis 1: There are patterns of development in collocational knowledge across proficiency levels

4.2.1.1. Essay Data (All Groups)

4.2.1.1.1. Kruskal-Wallis Analyses for the Essay Data

4.2.1.1.2. Summary of the Results for the Essay Data

4.2.1.1.3. Implicational Scaling for the Essay Data (All Groups)

4.2.1.2. Translation Data (All Groups)

4.2.1.2.1. Kruskal-Wallis Analyses for the Translation Data

4.2.1.2.2. Summary of the Results for the Translation Data

4.2.1.2.3. Implicational Scaling for the Translation Data (All Groups)

4.2.1.3. Blank-Filling Data (All Groups)

4.2.1.3.1. Kruskal-Wallis Analyses for the Blank-Filling Data

4.2.1.3.2. Summary of the Results for the Blank-Filling Data

4.2.1.3.3. Implicational Scaling for the Blank-Filling Data (All Groups)

4.2.1.4. Summary of the Results for Hypothesis 1

4.2.2. Hypothesis 2: There are patterns in the development of collocational knowledge within proficiency groups
4.2.2.1. Essay Data.................................................................162

4.2.2.1.1. Friedman Test for the Essay Data -
Group 1...............................................................................162

4.2.2.1.2. Friedman Test for the Essay Data -
Group 2...............................................................................163

4.2.2.1.3. Friedman Test for the Essay Data -
Group 3...............................................................................163

4.2.2.1.4. Implicational Scaling for the Essay Data by Groups..................167

4.2.2.1.5. Summary of the Results for the Essay Data........................170

4.2.2.1.6. Further Analyses on the Essay Data..............................170

4.2.2.2. Translation Data...............................................................171

4.2.2.2.1. Friedman Test for the Translation Data -
Group 1...............................................................................172

4.2.2.2.2. Friedman Test for the Translation Data -
Group 2...............................................................................173

4.2.2.2.3. Friedman Test for the Translation Data -
Group 3...............................................................................174

4.2.2.2.4. Implicational Scaling for the Translation Data by Groups..............175

4.2.2.2.5. Summary of the Results for the Translation Data.........................177

4.2.2.3. Blank-Filling Data..............................................................178

4.2.2.3.1. Friedman Test for the Blank-Filling Data -
CHAPTER 5: DISCUSSION OF THE FINDINGS

5.0. Introduction..........................188

5.1. Free Production of Collocations...............................................................189

5.1.1. Between-Group Differences..........................................................189

5.1.2. Within-Group Differences............................................................195

5.2. Cued Production of Collocations.............................................................200

5.2.1. Translation Data.................................................................200

5.2.1.1. Between-Group Differences.............................................201

5.2.1.1. Within-Group Differences...............................................205

5.2.2. Blank-Filling Data...............................................................211

5.2.2.1. Between-Group Differences.............................................212

5.2.2.2. Within-Group Differences...............................................214

5.3. Summary of the Findings.........................................................................221

5.4. Factors Affecting the Development of Collocational Knowledge....222
5.5. Summary of the Discussion..................................................................................229
5.6. Pedagogical Implications..................................................................................230

CHAPTER 6: CONCLUSIONS

6.1. Summary and Conclusions.............................................................................234
6.2. Directions for Further Research....................................................................237

APPENDIX A: Background information for the subjects of the study..................243
APPENDIX B: Questionnaire.................................................................................256
APPENDIX C: Essay topics....................................................................................258
APPENDIX D: Translation test.............................................................................261
APPENDIX E: Blank-Filling test...........................................................................263
APPENDIX F: Instructions given in the test..........................................................273
APPENDIX G: Teachers' questionnaire.................................................................274
APPENDIX H: Holistic rating scale in Jacobs et al. (1981)....................................277
APPENDIX I: Instructions for measuring Target-Like Use of Articles..................278
APPENDIX J: Instructions for measuring Lexical Density.................................280
APPENDIX K: Instructions for measuring T-Units and Error-Free T-Units.............283
APPENDIX L: Instructions for measuring Sentence-Nodes...................................284
APPENDIX M: Implicational scale for the translation data....................................284
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Examples of lexical functions</td>
<td>8</td>
</tr>
<tr>
<td>2. Examples of structural patterns of restricted collocations in English</td>
<td>43</td>
</tr>
<tr>
<td>3. Examples of grammatical and lexical collocations in the BBI Combinatory Dictionary of English</td>
<td>45</td>
</tr>
<tr>
<td>4. Types of collocation used in the study</td>
<td>50</td>
</tr>
<tr>
<td>5. Types of lexical collocations used in the study</td>
<td>52</td>
</tr>
<tr>
<td>6. The acquisition of English morphemes</td>
<td>57</td>
</tr>
<tr>
<td>7. Collocation tokens and token-type ratios in the TWE series</td>
<td>91</td>
</tr>
<tr>
<td>8. Means and standard deviations per textbook</td>
<td>93</td>
</tr>
<tr>
<td>9. Collocation types included in the translation test</td>
<td>98</td>
</tr>
<tr>
<td>10. Collocation types included in the blank-filling test</td>
<td>99</td>
</tr>
<tr>
<td>11. Kurtosis and skewness for the language proficiency measures</td>
<td>120</td>
</tr>
<tr>
<td>12. Means and standard deviations for the dependent variable: Holistic Rating</td>
<td>121</td>
</tr>
<tr>
<td>13. Post-hoc comparisons for the dependent variable: Target-Like Use of Articles</td>
<td>122</td>
</tr>
<tr>
<td>14. Means and standard deviations for the dependent variable: Target-Like Use of Articles</td>
<td>122</td>
</tr>
<tr>
<td>15. Post-hoc comparisons for the dependent variable:</td>
<td>xiv</td>
</tr>
</tbody>
</table>
Lexical Density

16. Means and standard deviations for the dependent variable:
   Lexical Density

17. Post-hoc comparisons for the dependent variable:
   Words per T-Unit

18. Means and standard deviations for the dependent variable:
   Words per T-Unit

19. Post-hoc comparisons for the dependent variable:
   Error-Free T-Units

20. Means and standard deviations for the dependent variable:
   Error-Free T-Units

21. Post-hoc comparisons for the dependent variable:
   S-Nodes per T-Unit

22. Means and standard deviations for the dependent variable:
   S-Nodes per T-Unit

23. Percentage of Error-Free T-Units in Larsen-Freeman (1978)

24. Summary of the results of the Kruskal-Wallis and post-hoc analyses for the essay data

25. Means and standard deviations by group for the essay data

26. Collocational use distinguishing proficiency levels

27. Summary of the results of the Kruskal-Wallis tests and post-hoc analyses for the translation data

28. Means and standard deviations by group for the translation data

29. Translation accuracy distinguishing proficiency levels
Summary of the results of the Kruskal-Wallis tests and post-hoc analyses for the blank-filling data .................................................. 153

Means and standard deviations by group for the blank-filling data .......................................................................................... 157

Blank-filling accuracy distinguishing proficiency levels .................. 159

Nemenyi's multiple comparisons tests of mean rank differences for the essay data - Group 1 .......................................................... 164

Nemenyi's multiple comparisons tests of mean rank differences for the essay data - Group 2 .......................................................... 165

Nemenyi's multiple comparisons tests of mean rank differences for the essay data - Group 3 .......................................................... 166

Nemenyi's multiple comparisons tests of mean rank differences for the translation data - Group 1 ...................................................... 172

Nemenyi's multiple comparisons tests of mean rank differences for the translation data - Group 2 ...................................................... 173

Nemenyi's multiple comparisons tests of mean rank differences for the translation data - Group 3 ...................................................... 174

Nemenyi's multiple comparisons tests of mean rank differences for the blank-filling data - Group 1 ...................................................... 181

Nemenyi's multiple comparisons tests of mean rank differences for the blank-filling data - Group 2 ...................................................... 182

Nemenyi's multiple comparisons tests of mean rank differences for the blank-filling data - Group 3 ...................................................... 183

Collocation types used most frequently in the students' essays........ 195
43. Accuracy order for the collocation types included in the translation test - All Groups........................................................................202

44. Accuracy order for the collocation types included in the translation test - Group 1....................................................................206

45. Accuracy order for the collocation types included in the translation test - Group 2....................................................................207

46. Accuracy order for the collocation types included in the translation test - Group 3....................................................................209

47. Accuracy order for the collocation types included in the blank-filling test - Group 1...............................................................214

48. Accuracy order for the collocation types included in the blank-filling test - Group 2...............................................................216

49. Accuracy order for the collocation types included in the blank-filling test - Group 3...............................................................217
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Distribution of collocations across the TWE series</td>
<td>91</td>
</tr>
<tr>
<td>2. Means use of collocation tokens - essay data - all groups</td>
<td>142</td>
</tr>
<tr>
<td>3. Mean use of collocation tokens per type in the essay data</td>
<td>146</td>
</tr>
<tr>
<td>4. Mean accuracy of response for the translation data</td>
<td>150</td>
</tr>
<tr>
<td>5. Mean accuracy of response for the translation data - all groups</td>
<td>152</td>
</tr>
<tr>
<td>6. Mean accuracy of response for the blank-filling data</td>
<td>158</td>
</tr>
<tr>
<td>7. Mean accuracy of response for the blank-filling data - all groups</td>
<td>161</td>
</tr>
<tr>
<td>8. Mean use of collocation tokens in the essay data - Group 1</td>
<td>168</td>
</tr>
<tr>
<td>9. Mean use of collocation tokens in the essay data - Group 2</td>
<td>169</td>
</tr>
<tr>
<td>10. Mean use of collocation tokens in the essay data - Group 3</td>
<td>169</td>
</tr>
<tr>
<td>11. Mean accuracy of response for the translation data - Group 1</td>
<td>176</td>
</tr>
<tr>
<td>12. Mean accuracy of response for the translation data - Group 2</td>
<td>177</td>
</tr>
<tr>
<td>13. Mean accuracy of response for the translation data - Group 3</td>
<td>177</td>
</tr>
<tr>
<td>14. Mean accuracy of response for the blank-filling data - Group 1</td>
<td>185</td>
</tr>
<tr>
<td>15. Mean accuracy of response for the blank-filling data - Group 2</td>
<td>185</td>
</tr>
<tr>
<td>16. Mean accuracy of response for the blank-filling data - Group 3</td>
<td>186</td>
</tr>
<tr>
<td>17. Model of the development of collocational knowledge</td>
<td>225</td>
</tr>
<tr>
<td>18. Mean age for subjects per Group</td>
<td>243</td>
</tr>
<tr>
<td>19. Distribution of gender across Groups</td>
<td>243</td>
</tr>
</tbody>
</table>
ABBREVIATIONS

Adj = Adjective
Adv = Adverb
BSM = Bilingual Syntax Method
c = Complement
crea(t) = Creation
det = Determiner
Diff = Difference
EFL = English as a Foreign Language
EFT = Error-Free T-Units
ELT = English Language Teaching
erad = Eradication
ESL = English as a Second Language
Hol. = Holistic
inf = Infinitive
L1 = Mother Tongue, First Language
L2 = Second Language
LD = Lexical Density
lex = Lexical
N = Noun
N = Node (only in Chapter 1)
NP = Noun Phrase
O = Object
poss = Possessive
Pred = Predicate
Prep = Preposition
S = Subject
S-Node(s) = Sentence Node(s)
STD DEV = Standard Deviation
TLU = Target Like Use
T-Unit(s) = Terminal Unit(s)
TWE = Task Way English
V = Verb
VP = Verb Phrase
wh = Wh-word
Acknowledgement is due to my supervisors for their help and support throughout my Ph.D. project. In particular, I thank Dr Michael Harrington for his patient editing of the drafts of chapters, and for his comments on the results and my interpretation of the findings, Dr Peter Robinson for helping me shape my research questions and analyse my data, Professor Roly Sussex for his help on the development of the research design and for discussions which were the starting point for my interest in collocations. I am also grateful to the students of the MA programme at C.L.T.R. for acting as inter-raters in the scoring of my data. Thanks are also due to Dr Toni Baglioni for his advice on the statistical analyses of my data.

I must also acknowledge the Principal of the ‘Fili;ppeion’ High School, Mr Michael Akrivopoulos, for permitting me to conduct my research at his school. Thanks are also due to the Vice-Principal, Mrs Sofia Pistofidou, for organising the data collection and for her support in obtaining permission for my data collection from the Department of Secondary Education in Greece.

I must also express my thanks to my family - my mother Maria, my father Dimos, and my sister Efi - for supporting and encouraging me throughout my study. I am also grateful to my life-long friend, Christina Bademi, for her support and assistance on numerous occasions. I would also like to express my gratitude to my friends, Professor Andrew Lian and Ania Lian for their support and friendship over the past 4 years. Most of all, though, I have relied on the help and support of Richard Taylor, who kept me happy.
while I wrote this, sharing the many labours with me. I am more grateful to him than I can say.

Finally, I must thank DEET for supplying me with an Overseas Postgraduate Scholarship that paid for my tuition fees, and C.L.T.R. for employing me throughout the majority of my time in the Ph.D. programme, thus allowing me to support myself financially.
To my Parents

Afierwvvnw th diatribhv mou stouß Goneivß mou,
ston patevra mou, Dhvmo Gkitsavkh,
kai sth mhtevra mou, Mariva Liavtsou-Gkitsavkh,
me agavph.