On Second Language Learner Acquisition of English Collocations

James Martin Rogers
Meijo University, Japan

Series in Education
For my daughter
# Table of Contents

List of Figures and Tables v

Preface xi

Acknowledgments xiii

Chapter 1 **Introduction** 1

Chapter 2 **Review of the Literature** 5

Introduction

Overview of Second Language Learning Research

On Learners’ Lack of Collocational Fluency

On the Learning Burden of Collocations

On Approaching and Defining Collocations

  What is a Collocation?

  Types vs. Lemmas vs. Word Families

  On Semantic Transparency

  On Concgramming, MWU Length, and Colligation

On the Value of Collocations

On the Lack of Research and Resources

On the Direct Teaching of Collocations

On Utilizing Corpus Data to Identify Collocations

On Using L1-L2 Congruency to Identify Useful Collocations

Conclusion

Chapter 3 **Addressing Gaps in the Research** 37

Introduction

The Gaps in the Research
Research Methods and Techniques

Research Paradigm

Data Source, Collection Methodology, and Analysis

Filling the Gaps in the Research

What is an Ideal Corpus Frequency Data Cut-off for Identifying High-frequency General English MWUs?

Is Corpus Dispersion Data Reliable for Identifying High-frequency General English MWUs?

Is Corpus Chronological Data Reliable for Identifying High-frequency General English MWUs?

Is Consideration for Colligation an Important Criterion for Identifying High-frequency General English MWUs?

What Percentage of High-frequency General English MWUs is Deemed by Fluent Speakers Worthy of Expanding Beyond Their Most Frequent Exemplar?

What Percentage of MWUs Most Representative of High-frequency General English Lemmatized Concgrams Has Low Semantic Transparency?

What Percentage of MWUs Most Representative of High-frequency General English Lemmatized Concgrams Has Low L1-L2 Congruency With Japanese, Chinese, and Korean?

Is Fluent Speaker Intuition Reliable Regarding High-frequency Vocabulary Usage in Context Creation?

What is Japanese University Students’ Knowledge of MWUs Most Representative of High-frequency General English Lemmatized Concgrams?

Conclusion

Chapter 4  **Implications and Applications**  87

Chapter 5  **Conclusion**  101

References  105

Index  115
List of Figures and Tables

List of Figures

Figure 3.1. Percentage of Items Accurately and Erroneously Flagged for Balanced Dispersion Data Distribution at All Three Parameters 48

Figure 3.2. Total Items Erroneously Flagged or Judged Unbalanced Which Were Not Flagged 48

Figure 3.3. Percentage of Items Accurately and Erroneously Flagged for Balanced Chronological Data Distribution at all Three Parameters 55

List of Tables

Table 2.1. The Three Approaches’ Ability to Identify Common Collocates of the Verb ‘Play’ 13

Table 2.2. High-frequency Collocations for the Four Most Frequent Words in the Word Family for ‘Govern’ According to the COCA (Top Frequencies in Bold) 16

Table 2.3. A Sample of Data From the COCA for a Concgram Search for the Lemmas ‘Provide’ and ‘Support’ 19

Table 2.4. Top Three MWUs for the Lemmas ‘Provide’ and ‘Support’ Found After Examining 500 Concordance Strings in the COCA 20

Table 2.5. MWUs Identified From 500 Example Sentences in Which the Lemmas ‘Come’ and ‘Term’ Both Occur in 21

Table 2.6. A Comparison Between Two MWU Searches, One With and One Without Consideration for a Specific Type of Colligation 22

Table 3.1. Word Frequency Breakdown of Lemma Pairs Occurring Once Per Million Tokens According to Vocabprofile’s 25,000-word Families of the BNC and COCA 44

Table 3.2. Criteria for Rating the Value of Collocates for Learners of General English 47

Table 3.3. Items Found Not to be Worthy of Inclusion Because They Were Either Inappropriate Language, Grammatical Formulations, Duplicates, or Compound Nouns 49
Table 3.19. Word Family Frequency Breakdown of Formulaic Phrases Within Example Sentences Created Using Fluent Speaker Intuition Using RANGE

Table 3.20. Vocabprofiler Breakdown of Top 3,000 Word Family Words Not Covered by Example Sentences Created Using Fluent Speaker Intuition

Table 3.21. L1-L2 Congruency Ratings of MWUs Selected for Testing Students' Collocational Fluency

Table 3.22. Mean Scores for Test Items Organized by Frequency Level

Table 3.23. Multiple Regression Analysis and Correlation Coefficient with TOEFL as the Dependent Variable and Item Frequency as the Independent Variable

Table 3.24. Multiple Regression Analysis and Correlation Coefficient with TOEFL Score as the Dependent Variable and L1-L2 Congruency as the Independent Variable

Table 4.1. Top Ten TOEIC Gains in Comparison to Minutes Studied with the App
Preface

Gaining control of collocations is a very important part of language learning, but learners are typically slow in sounding native-like in their use of collocations. One of the reasons for this is that the majority of collocations occur infrequently. It is useful to bear in mind that the frequency of any collocation will be lower than the frequency of its lowest-frequency member.

Collocations have the chance of being learned across the four strands of a well-balanced course. In the meaning-focused input strand of a course, collocations will be met through listening and reading. Because learning through the strand of meaning-focused input is largely incidental learning, large quantities of input through extensive reading and extensive listening are needed to get enough repetitions of particular collocations for them to stay in memory. Collocations also have the chance to be learned through meaning-focused output, that is, through speaking and writing, where learners have the opportunity to turn receptive knowledge into productive use and to strengthen and enrich their knowledge of collocations. The fluency development strand is likely to be a very important strand in the development of collocational knowledge as the pressure to use language faster encourages a focus on restructuring knowledge to achieve fluency by moving from the processing of single words to greater use of multiword units. These three meaning-focused strands all depend on incidental learning and frequency of occurrence.

The strand of language-focused learning provides the opportunity to speed up the learning of collocations through deliberate teaching and deliberate learning. Such deliberate teaching and learning are most efficiently done if there are well-researched lists of collocations to draw on. This book reports on research that has been done with this goal in mind.

The case for giving deliberate attention to collocations is strengthened by research that suggests that, as with the deliberate learning of vocabulary, the deliberate learning of collocations simultaneously results in both implicit knowledge and explicit knowledge. That is, the deliberate learning of collocations will provide the kind of knowledge that is needed for normal language use.

In order for something to be learned, there are three requirements of attention. Firstly, what needs to be learned needs to be focused on. That is, the form, the meaning, and/or the use of the collocation needs to be focused on. Secondly, there needs to be a quantity of attention. In other words, there
needs to be repeated encounters and good attention for each encounter. The greater the amount of attention, the more likely the collocation is to be remembered. Thirdly, there needs to be good quality attention. In the case of collocation, the quality of attention can be increased by looking at how the parts of the collocation relate to the meaning of the whole, the use of the collocation in a variety of contexts, and visualizing the meaning of the collocation. These requirements of attention apply to both incidental and deliberate learning. In incidental learning, the three requirements are typically not as strongly applied as in deliberate learning. In incidental learning of collocations, as in listening and reading, there tends to be brief attention to the form, with similarly brief meaning recall. Each incidental meeting tends to involve a small amount of attention, but with repeated meetings, there can be a cumulative effect. The quality of attention is typically not very deep, involving a brief focus on the form and contextual meaning.

The research in this book provides very useful support for the deliberate learning of collocations. It makes sure that, in deliberate learning, attention is focused on the most useful items to learn.

Paul Nation
Acknowledgments

I would like to thank all research assistants and translators who volunteered their time for my research projects in the past. Without their help, I wouldn't have been able to tackle the large-scale research projects this book will describe. Filling the gaps in the research would have also been difficult to accomplish if not for the software solution that was provided by Dr. Laurence Anthony.
PAGES MISSING
FROM THIS FREE SAMPLE
References


Byrd, P., & Coxhead, A. (2010). ‘On the other hand’: Lexical bundles in academic writing and in the teaching of EAP. University of Sydney Papers in TESOL, 5, 31-64.


References


References


References


<table>
<thead>
<tr>
<th>C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>chronological</td>
<td>30, 34, 38, 40, 41, 45-47,</td>
</tr>
<tr>
<td></td>
<td>54-57, 84, 88-90, 102, 104</td>
</tr>
<tr>
<td>colligation</td>
<td>18, 21-23, 33, 38, 39, 57,</td>
</tr>
<tr>
<td></td>
<td>59, 64, 65, 85, 95</td>
</tr>
<tr>
<td>colligational</td>
<td>23, 41, 58-65, 89, 95, 103</td>
</tr>
<tr>
<td>concgramming</td>
<td>2-4, 20, 26, 32</td>
</tr>
<tr>
<td>concgram(s)</td>
<td>18-20, 32, 38, 42, 46, 54,</td>
</tr>
<tr>
<td></td>
<td>57, 62-68, 70, 71, 73, 78-80,</td>
</tr>
<tr>
<td></td>
<td>83-86, 89-92</td>
</tr>
<tr>
<td>congruency</td>
<td>2, 5, 10, 11, 17, 18, 26,</td>
</tr>
<tr>
<td></td>
<td>31, 32, 34, 37, 38, 41, 42,</td>
</tr>
<tr>
<td></td>
<td>45, 69-73, 78-80, 82-86, 93,</td>
</tr>
<tr>
<td></td>
<td>96, 97, 102</td>
</tr>
<tr>
<td>constituent variation</td>
<td>2, 18, 20</td>
</tr>
<tr>
<td>core idiom(s)</td>
<td>18, 42, 68, 69, 85, 92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>dispersion</td>
<td>30, 34, 38, 40, 41, 45-49,</td>
</tr>
<tr>
<td></td>
<td>51-57, 84, 88, 89, 102, 104</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>figurative(s)</td>
<td>17, 18, 42, 68, 69, 85,</td>
</tr>
<tr>
<td></td>
<td>88, 92, 95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>literal(s)</td>
<td>9, 17, 18, 33, 42, 68-71,</td>
</tr>
<tr>
<td></td>
<td>85, 88, 92, 95, 96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>11, 29, 30, 34, 40</td>
</tr>
<tr>
<td>mutual information</td>
<td>11</td>
</tr>
</tbody>
</table>