

NON-NATIVE SPEAKER COLLOCATIONS: A CORPUS-DRIVEN
CHARACTERIZATION FROM THE WRITING OF
NATIVE SPEAKERS OF MANDARIN

by

ROBIN JANINE LOMBARD

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 1997

Copyright © by Robin Janine Lombard 1997

All Rights Reserved

ACKNOWLEDGMENTS

No one is able to successfully complete a dissertation project without advice, support, and encouragement; in that respect this dissertation is no different from any other. There were many, many people who contributed to the success of this endeavor. First and foremost, I would like to thank my supervising professor, Dr. John Paolillo, who worked with me from the very beginning, giving me what I needed most: time. From him I have learned not only how to complete a large research project, but what it means to be a mentor. I would also like to thank the other four committee members, Dr. Irwin Feigenbaum, Dr. Lana Rings, Dr. David Silva, and Dr. Pete Smith for their helpful comments and suggestions. Each one contributed to the clarity and accuracy of this document. Special thanks also goes to Li Wen-qi, who provided invaluable insight into differences in Mandarin and English, and Robert Johnson, who graciously helped me gather the data.

In addition, there were many friends and family members who supported me as I went through the difficult process of putting my thoughts on paper. I would like to thank all those people who were there on a daily basis to hear about my progress and encouraged me to continue working: Pam and Steve Echerd, Jo and David Lombard, Arlene Bolton, Chip, Debra Basler, Jenny Lonas, Carole Nix, Tac Tacelosky, Deborah Johnson, the faculty and staff of the English Language Institute at UT Arlington, the members of the Dream Team, the members of the Clan MacGab, and so many others. I wouldn't have made it through without each and every one of them.

November 26, 1997

ABSTRACT

NON-NATIVE SPEAKER COLLOCATIONS: A CORPUS-DRIVEN
CHARACTERIZATION FROM THE WRITING OF
NATIVE SPEAKERS OF MANDARIN

Publication No. _____

Robin Janine Lombard, Ph.D.

The University of Texas at Arlington, 1997

Supervising Professor: John C. Paolillo

Non-native speaker difficulty in producing collocations (i.e., domain-specific recurrent word combinations) is widely acknowledged (see, e.g., Bahns 1993, Channel 1981, Farghal and Obiedat 1995, Gitsaki 1996, and Zughoul 1991). Phrases such as **a convenient price* or **marketing share* instead of the native speaker collocations *a reasonable price* and *market share*, for example, comprise a substantial number of advanced learner errors written by native speakers of Mandarin in a business English context. This problem has been addressed in the form of domain-specific reference and teaching materials (see Flower and Martínez 1995 and Tuck 1993). While these materials—which are based on analyses of native speaker texts and/or native speaker intuition—accurately represent native-speaker language, it is unclear whether they meet the needs of the language learner. The actual collocational needs of the learner can only be confirmed through an analysis of "learner" collocations themselves.

This study examines idiosyncratic English collocations produced by native speakers of Mandarin in written assignments for an MBA English preparation course. The collocations themselves were identified, divided into types, and examined in a series of

qualitative and quantitative analyses. The objective of the study was to synthesize from these analyses a characterization of learner collocation.

The initial qualitative analysis identified four categories of idiosyncrasy in learner collocation: lexical transfer, and phonological, grammatical, and semantic influence. Those learner collocations classified as containing a semantic idiosyncrasy comprised more than half the total. Quantitative analyses revealed that learner collocations tend to involve an idiosyncrasy in either form or meaning, but rarely both. Learner collocations involving semantic idiosyncrasy were further analyzed into two parts: a headword and a functor. The site of the idiosyncrasy was most often the functor, and rarely occurred on both parts of the same collocation simultaneously.

This analysis of learner collocations provides insight into the nature of learners' distinctive production problems with collocation, and supports the creation of effective teaching materials and reference works based upon the difficulties demonstrated by the learner in producing collocations.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ABSTRACT	v
LIST OF ILLUSTRATIONS	x
LIST OF TABLES	xi
Chapter	
1. INTRODUCTION	1
1.0 Presentation of the problem	1
1.1 Implications of the problem	5
1.2 An approach to the problem	7
1.3 Chapter overview	10
2. REVIEW OF THE LITERATURE	11
2.0 Introduction	11
2.1 Linguistic/lexicographic definitions of collocation	13
2.2 Studying specific domains	26
2.3 English for specific purposes	27
2.4 Assumptions of this study	31
3. DATA AND METHODOLOGY	33
3.0 Introduction	33
3.1 The use of corpora in research	34
3.2 Data-Gathering Context	36
3.3 Participants	37

3.4	Data types	39
3.5	Data collection	45
3.6	Data processing: Gathering and storing the collocations	46
3.7	Coding	53
3.8	Conclusion	58
4.	RESULTS AND ANALYSIS	59
4.0	Introduction	59
4.1	Collocational idiosyncrasy involving primarily lexical transfer	60
4.2	Collocational idiosyncrasy involving primarily phonological similarity	68
4.3	Collocation idiosyncrasy involving primarily irregularities in grammar	76
4.4	Collocation idiosyncrasy involving semantic choice variation	85
4.5	Residue	103
4.6	Conclusions	103
5.	DISCUSSION OF THE FINDINGS	104
5.0	Introduction	104
5.1	Initial analysis of observed idiosyncratic collocations	105
5.2	Examination of idiosyncrasies involving grammar	125
5.3	Examination of idiosyncrasies involving semantics	129
5.4	A proposed characterization of learner collocations	145
6.	CONCLUSIONS	147
6.0	Summary and conclusions	147
6.1	Implications	148
6.2	Directions for further research	153
Appendix		
A	HUMAN SUBJECTS COMMITTEE APPROVAL LETTER AND CONSENT FORM	156

B	STUDENT QUESTIONNAIRES	159
C	READING TEXT BIBLIOGRAPHY	164
D	COLLOCATION CROSS-CLASSIFICATION FOR EACH STUDENT	166
E	CONTINGENCY TABLES OF COLLOCATION TOKENS FOR EACH STUDENT	177
F	CONTINGENCY TABLES OF COLLOCATION TYPES FOR EACH STUDENT	202
	REFERENCES	227

LIST OF ILLUSTRATIONS

Figure		Page
1.	Conc Concordance Fragment	47
2.	Collocation Database Record	52

LIST OF TABLES

Table	Page
1. Student Entry and Exit Scores	38
2. Student Background Information	40
3. Writing Tasks	42
4. Number of Words per Writing Task	43
5. Class Reading Assignments	45
6. Lexical Relations Used in this Study	57
7. Idiosyncratic Collocation Involving Direct Translation	61
8. Idiosyncratic Collocation Involving Word Matching	62
9. Idiosyncratic Collocation Involving Number Matching	63
10. Idiosyncratic Collocation Involving Primarily Lexical Transfer	64
11. Lexical Transfer Collocation Summary	65
12. Lexical Transfer Collocations per 100 words	67
13. Phone-level Idiosyncratic Collocation	69
14. Syllable-level Idiosyncratic Collocation	70
15. Word-level Idiosyncratic Collocation	71
16. Idiosyncratic Collocation Involving Primarily Phonology	72
17. Phonological Similarity Collocation Summary	74
18. Phonological Similarity Collocations per 100 words	75
19. Idiosyncratic Collocation Involving Word Form	77
20. Idiosyncratic Collocation Involving Preposition	78

21.	Idiosyncratic Collocation Involving Genitive Structure	79
22.	Idiosyncratic Collocation Involving Word Order	80
23.	Idiosyncratic Collocation Involving Mixed Variation	80
24.	Idiosyncratic Collocation Involving Primarily Grammar	81
25.	Grammar Collocation Summary	83
26.	Grammar Collocations per 100 words	84
27.	Semantic Choice Collocation Summary	98
28.	Semantic Choice Collocations per 100 words	99
29.	Total Collocation Summary	101
30.	Total Collocations per 100 Words	102
31.	Collocation Residue	103
32.	Grammar vs. Phonology Tokens Contingency	106
33.	Semantic vs. Lexical Transfer Tokens Contingency	107
34.	Phonology vs. Lexical Transfer Tokens Contingency	108
35.	Phonology vs. Semantic Tokens Contingency	109
36.	Grammar vs. Lexical Transfer Tokens Contingency	110
37.	Grammar vs. Semantic Tokens Contingency	111
38.	Token Chi-square Values for Each Student under Each Contingency	113
39.	Grammar vs. Phonology Types Contingency	114
40.	Semantic vs. Lexical Transfer Types Contingency	115
41.	Phonology vs. Lexical Transfer Types Contingency	115
42.	Phonology vs. Semantic Types Contingency	117
43.	Grammar vs. Lexical Transfer Types Contingency	118
44.	Grammar vs. Semantic Types Contingency	119
45.	Type Chi-square Values for Each Student under Each Contingency	121
46.	P-Value Summary for Token Contingencies	122

47.	Native Structure/Unknown Influence vs. Grammar Categories	127
48.	Presence of Headword and Functor Idiosyncrasy	131
49.	Headword Idiosyncrasy Versus Mel'čuk Reduced Categories.....	133
50.	Functor Idiosyncrasy Versus Mel'čuk Reduced Categories.....	134
51.	Bon Category with Headword/Functor Distinction	136
52.	Adj Category with Headword/Functor Distinction	138
53.	Mult Category With Headword/Functor Distinction	139
54.	Plus Category with Headword/Functor Distinction	140
55.	Magn Category with Headword/Functor Distinction	142
56.	Collocations Involving both Headword and Functor Idiosyncrasy	144

CHAPTER 1

INTRODUCTION

1.0 Presentation of the problem

The non-native speaker of English (hereafter NNS) who is in the U.S. to gain higher education in an academic specialty faces two daunting tasks: to absorb and process specialized information in a foreign language, and to learn to produce native-sounding language in a specific language domain. Of these two, the latter is the more difficult. One of the primary issues for such advanced learners involves the correct use of vocabulary and phraseology characteristic of a given language domain. Specifically, NNS problems in the written production of sets of word combinations which native speakers formulate naturally in text are well documented (see, e.g., Palmer 1933, Channel 1981, Brown 1988, Bahns 1993, Farghal and Obiedat 1995, and Gitsaki 1996). For example, in a memo to his superior, a NNS engineer wrote **the design of the logics* in place of the conventional engineering phrase, *logical designs*. A NNS businessman in a formal letter mentioned in closing that he hoped the recipient would respond in **the soon future* rather than the accepted *near future*. And a NNS business student wrote that her proposed product would **increase the quality of life*, instead of *improving* it. The word combinations exemplified above—*logical designs*, *near future*, *improve the quality of life*—are commonly referred to in the literature as "collocations."

That collocations are of vital importance to restricted language domains is well documented (Smadja 1989, 1993b). While meaning can be negotiated in spoken contexts, written contexts necessarily require more precision in wording. Therefore, advanced

learners of English who are seeking to become part of the conversation in restricted language domains must be able to produce in written form collocations prevalent in those domains (see, e.g., Kjellmer 1987 and Knowles 1993). The graduate student in business, for example, needs to be capable of producing the collocations specific to the graduate business English domain in order to gain membership in that discourse community.

The "learner" collocations—idiosyncratic phrases approximating native speaker produced collocations—shown in examples 1 through 4 are typical of those found in the writing of native speakers of Mandarin. Each of the four examples illustrates a different type of problem in the production of collocations (in bold type) which mark the writer as non-standard.

Example 1

- a) People just stay at home, **open the monitor and computer**, and connect central control institute.
- b) For instance, the Federal Home Loan Board (FHLB), the most influential agency in the United States, evaluates members, whether or not, can **borrow urgent funds** from the central bank.

In example 1, the problem is one of direct translation of a lexical item from Mandarin into English. A native speaker would have written *turn on the monitor and computer* and *borrow emergency funds*, but the native speaker of Mandarin in each case chose one word which, though appropriate in Mandarin, doesn't fit the native English speaker equivalent of the collocation.

Example 2

- a) Although a trust is a taxable entity, individuals can use a trust to shift their income to lower **tax rate basket** since trust property is separate from the settor's taxable assets.
- b) However, the company now encounters the **intensive competition**, which resulting in the drastic decrease of its marketshare.

Example 2 represents phrases in which a NNS chooses a word phonologically (and perhaps semantically) very close to the word a native speaker would most likely produce. The phrase **intensive competition* would be rendered by a native speaker as *intense competition* and **tax rate basket* would be *tax bracket*.

Example 3

- a) ... the difference of proportions between targeted markets and real **consumers of women** is 20% less than that of men ...
- b) The **key of the good strategies** in overseas investment is selecting the right countries and the correct investing composition.

The problem depicted in examples 3a and 3b involves structural idiosyncrasy; the phrases used by a native speaker in these examples would be *women consumers* and *the key to good strategies*.

Example 4

- a) As a result of the financial status and **customer psychology** of students, the price strategy should be reasonable but higher than the similar products for students.
- b) Lifestyles emerge from various social influences and are also derived from the individuals' personal **value concepts** and personalities.

The issue exemplified in 4a and 4b is one of semantic idiosyncrasy of the use of *customer* and *concepts*; native speakers would have written the words *consumer* and *judgements*, respectively.

Much has been reported in the literature about the influence of native language on the construction of learner collocations (see, e.g., Bahns 1993, Farghal and Obiedat 1995, Hussein 1990, and Zughoul 1991). Many researchers attribute the greatest influence in the

construction of learner collocations to the native language. Certainly, evidence of native language transfer is found in this study, but not all learner collocations can be accounted for by direct lexical transfer from the native language. Three other plausible influences on the production of idiosyncratic collocations are identified in this study: phonology, grammar, and semantics. The phonological influence appears in all parts of words, but most especially in the final syllable. For example, *reverse this drawback* becomes **reserve this drawback*, and *relatively weak* becomes **relevant weak*. Grammar influences the structural elements of learner collocations. Genitive structures (**lives' quality*), prepositions (**attention is paid on them*), word form (**food, clothes, and shelters*), and word order (**more worth*) all contribute to the grammatical category of idiosyncratic collocations. Learner collocations containing a semantic idiosyncrasy (e.g., **translating the dollars*, **work out poverty*, and **middle level persons*) are the most prevalent type found in this study.

Collocations are difficult constructions for the NNS to produce for several reasons. Though they are clearly word groups, collocations are not parsed any differently than other phrases. That is, their lexical "unithood" is not marked phonologically, syntactically, or orthographically, so they do not require that their special status be recognized, as would be the case with idioms. In other words, when either a native speaker or a NNS reads or hears a collocation, he or she does not need to recognize it as a unit; the meaning of each word (and the entire phrase) is clear if one follows normal parsing rules. It is only when the time comes to produce a collocation that certain constraints come into play, constraints that one must follow but has difficulty explaining to NNSs because they are completely unmarked.

An explanation of why collocations are difficult for the NNS to master is much easier to provide than an explanation of the learner collocations NNSs do produce. All four categories of errors in word choice and phrasing in the examples above are quite common even in the writing of NNSs who are at an advanced stage in their acquisition of English.

Though attested in second language acquisition research, NNS problems in the production of English collocations, such as those presented here, have been neither rigorously defined nor thoroughly explored.

1.1 Implications of the problem

Published studies on collocation as it pertains to learners of English tend to focus on two points: that collocations are a problem for learners of English, and that there is a need for teaching and reference material geared specifically toward collocation. In fact, the need is well enough recognized that there now exist reference materials for learners which include collocation information. One such reference is Tuck's *Oxford Dictionary of Business English* (1993), designed specifically for learners of English and based on "the combined expertise of people in the world of business and those involved in the teaching of English as a foreign language (iii)." This dictionary contains a special section in each entry listing common collocations; it is not clear whether these collocations were gathered from an analysis of native-speaker texts or from human intuition, or both.

At first glance, this specialized reference work would appear to fill the need evidenced above. However, of the eight learner collocations presented in examples 1 through 4, only one of the corresponding native speaker collocations appears in the *Oxford Dictionary of Business English*: *tax bracket* (easily found under the word *tax*). On the other hand, the collocation *intense competition* could not be found under entries for either word. Of the remaining six collocations, four (*turn on the monitor*, *borrow emergency funds*, *consumer psychology*, and *women consumers*) had one member appear in the dictionary in other collocations, but not in exactly the same collocation as that approximated by the learner. The last two collocations (*key to good strategies* and *value judgement*) also each had one member of the collocation appear in the dictionary, but the sense of the word was different from the student's intended use. The lack of a match between learner

collocations and native-speaker equivalents in the dictionary indicates that this dictionary might not be as useful for this group of students as it is intended to be, simply because those collocations chosen for inclusion were not the ones students had trouble with.

A large part of the problem in teaching NNSs about collocation may be that though fairly well understood from the native speaker standpoint, only a limited understanding of the nature of learner collocations actually informs teaching and reference materials. The illustration presented in this section of how a specialized dictionary failed to meet learner needs for collocational information demonstrates how important such an understanding of the nature of learner collocation could be. In order to fully understand how or why the NNS fails to produce native-sounding collocations, irregular versions of collocations must be identified, examined, and categorized.

Collocation, unlike other facets of language instruction, is a language feature that teachers of NNSs themselves may only be subconsciously aware of. A problem on the word level, for example, is addressed by the teacher who defines unfamiliar words, or the student who refers to the dictionary for meaning. Teachers present phrase structures in paradigm fashion, allowing students to quickly learn restrictions on replacement items. Idioms, though more complex than words, are learned in much the same way; they are easily identified because of their markedness and in fact cannot even be processed by the NNS without some form of assistance. In contrast to these elements of language, the collocation is a stumbling block for the NNS precisely because of its lack of markedness, which makes it difficult to predict, identify, and teach.

A dictionary aimed at NNS problems with collocations in restricted language domains is a sound idea for meeting the needs of the advanced learner. However, the design of the dictionary must take into account the perceptions and error patterns of the NNS. Therefore, the collocations chosen for inclusion in a specialized dictionary should be based at least in part on what students actually struggle with, not just collocations that are

the most frequent, or those that are judged by means of native-speaker intuition to be most important.

1.2 An approach to the problem

The question requiring an answer centers around the nature of the learner collocation. Gitsaki's (1996) study of the acquisition of collocations by NNSs under very controlled conditions determined that there is a specific order of acquisition in collocations. She claims that the more complex and rarely used collocations are extremely difficult for the NNS to acquire, while simple, frequently observed collocations are those most likely to be accurately used by the NNS. If Gitsaki's results are sound, they support the hypothesis that it is the idiosyncratic, irregular collocations produced by the NNS that need to be studied. In fact, Gitsaki's results have led her to suggest that collocational errors themselves need to be examined further. What is needed, then, is a characterization of learner collocations. This approach to the problem of collocations and NNSs requires that several factors be taken into consideration.

The first aspect of a study of learner collocations which must be addressed is their collection: how can they be collected for analysis? Many recently published descriptions of collocations and investigations into the nature of collocation employ a quantitative corpus analysis approach, in which large quantities of naturally occurring (native-speaker produced) texts are compiled and analyzed to uncover statistical patterns in language use. The corpus approach has advantages for studying learner language as well, and is an approach endorsed by Granger (1992), whose studies on learner language involve the use of learner language corpora. The collection of a corpus of learner language allows the researcher access to language in context, and provides data that may be significantly different from specific elicitation tasks, such as the blank filling task used in Farghal and Obiedat (1995), Gitsaki (1996), and Hussein (1990) in which NNSs are given part of a

collocation and must supply the rest. The corpus analysis approach is therefore used in the present study of learner collocation in order to ensure both that what is being studied is what students have themselves chosen to produce, and that the larger discourse context is retrievable.

A corpus is almost always limited by certain factors such as text type, number of texts, depth of each text, and so forth, depending on the nature of the given study. A corpus-driven study of learner collocation is most easily performed with a well-focused corpus in terms of the learners and their context. Therefore, a specific genre (business English), mode (written texts), native language (Mandarin), and context (classroom assignments) are all factors in the creation of the corpus used in this study.

While a corpus of student texts provides a context, it also presents a problem: how will learner collocations be identified? This is a difficult question, one which this study addresses by examining the definitions of collocation present in the literature of various disciplines. The answer, finally, involves an innovative definition specific to the collocations produced by NNSs, based on the assumption that while native speakers cannot necessarily identify a collocation in running text they are quite adept at identifying collocational violations. Therefore, instead of relying upon a specific definition previously put forward, a set of characteristics gleaned from those definitions is proposed to help identify NNS collocations. Learner collocations are thus identified as word groupings in a syntagmatic relationship which display a non-grammatical semantic constraint in at least one of the members, together embody a meaning recoverable from the component parts, do not allow for replacement items, and are seen by the native speaker as a lexical unit.

The identification of learner collocations is not the end of the process, however. Once they have been identified, they must be classified: according to what criteria can learner collocations be grouped? The first possibility explored in this study was language transfer. One method of identifying cases of direct lexical transfer from the native language

is the use of back-translations. This process involves providing a person who speaks both languages with the English version of the collocation, which the bilingual then translates into the other language (in this case, Mandarin). The foreign language equivalent of the English collocation is then back-translated word-for-word into English, and is then compared with learner collocations to locate cases of direct lexical transfer.

Since not all learner collocations involve lexical transfer, other possible classifications have to be explored. Dechert and Lennon (1989) report the production of English collocation "blends" by native speakers of German, in which the NNSs juxtapose parts of two different English collocations. That finding was rare in this study, and for that reason was not pursued after other more dominant categories became apparent. Instead, learner collocations herein that did not involve lexical transfer were classified as involving either phonological, grammatical, or semantic idiosyncrasy. In addition, learner collocations were cross-classified according to possible secondary influences.

Additional division into sub-categories allowed for a more fine-grained analysis. Learner collocations involving lexical transfer were found to be of two types: those involving direct translation and those involving a mismatch between words in Mandarin and English. Phonological idiosyncrasy occurred at all levels (phone, syllable, and word) and in different locations inside the word (initial, medial, and final syllables). Collocational idiosyncrasy as a result of grammar structure difficulties includes five types: word form, word order, genitive, preposition, and mixed. Those learner collocations classified as involving a semantic idiosyncrasy were further divided according to twenty of Mel'čuk's (1986) lexical relations.

Once the learner collocations have been identified and classified, they can be further analyzed according to quantitative methods. In this study, quantitative analysis included the extensive use of the chi-square test to determine the probability that the four main categories were operating independently. Additionally, learner collocations involving semantic

idiosyncrasy were divided into headword and functor; results of analyses on these also contributed to the overall characterization of learner collocations.

The goal of the present study, then, is to generate a characterization of learner collocations written by native speakers of Mandarin through a corpus analysis of their classroom writing.

1.3 Chapter overview

This dissertation is organized as follows: chapter 2 presents a review of the literature, with special attention paid to collocation in general; the chapter closes with a presentation of the definition of collocation used in the identification of learner collocations in the corpus. The context of the study, data types, and the construction of the corpus are outlined in chapter 3. Methods employed in gathering the data for the corpus, identifying learner collocations, and storing and analyzing the learner collocations are included here as well. Chapter 4 describes the object of study, the collocations themselves, and puts forward a first qualitative analysis of the learner collocations based on the nature of their idiosyncrasies. Chapter 5 continues with further analysis and discussion of the collocations based on various qualitative and quantitative investigations, and concludes with a tentative characterization of learner collocations produced by native speakers of Mandarin. Chapter 6, the final chapter, provides an overview of the results of the study, along with a discussion of its implications and directions for further research.

CHAPTER 2

REVIEW OF THE LITERATURE

2.0 Introduction

This chapter will analyze various definitions of collocation that have been put forth in the literature, specifically addressing two questions: How has collocation been defined?; and How can collocation best be defined for a specific study in the domain of English for Specific Purposes? Because so many definitions of collocation have been proposed in the literature, based on different theories of language, an examination of them is a vital precursor to this study. A survey of prevalent definitions reveals wide variation in meaning, as well as significant gaps in the literature for and by ESL/ESP professionals; even within sub-disciplines, there is no centrally agreed-upon concept. In addition, the issue of whether a general definition of collocation is suitable for a specific language domain must be examined. In order to construct a single definition of collocation that will be useful in limited language domain contexts, this chapter will examine the term itself as well as how it has been appropriated by researchers in various fields to represent rather different concepts related to word patterning in text. Through a comparison of these definitions and their underlying assumptions the elements to be included in the concept of collocation used in this study will be clarified, which will lead to the goal of this section: to propose a definition of collocation suitable for any specific language domain encompassed by English for Specific Purposes.

Defining what collocation means in this study is an essential first step, because of the many different definitions of collocation that already exist. Herbst (1996) attributes the

present variation in definitions to the varying interpretations given to Firth's original work on the subject (Firth 1957), which in Herbst's estimation is quite vague itself. He identifies three approaches commonly taken in the investigation of collocation: the text-based approach; the statistically oriented approach; and the significance oriented approach (380). These approaches, applied in various disciplines, have spawned definitions that in some cases bear little resemblance to one another.

Choosing a single existing definition implies choosing a set of assumptions about language. For example, existing definitions are based on either analyses of native speaker-produced text or native-speaker intuition. Since this study examines NNS data, the use of an existing definition may result in a failure to identify key characteristics of collocation because the data will be viewed within a paradigm that contains an underlying native-speaker bias. In other words, there may be factors present in the NNS understanding and construction of collocation that have not yet been taken into account by existing definitions. Moreover, many definitions, such as that proposed by Smadja (1993a), are discipline-specific. A suitable definition of collocation for this study must be able to account for the data. This means either working under a definition which exists in the ESL/ESP discipline, or constructing a new one.

Chapter 2 will begin by examining existing definitions of collocation put forward by linguists and lexicographers, and identifying characteristics and assumptions that are germane to this study (section 2.2). Section 2.3 briefly addresses language in specific domains. Section 2.4 will then analyze definitions appearing in the ESL/ESP literature, demonstrating how they fail to fully account for the language data presented here. This overview attempts to show that, for a study of this nature, insufficient work has been done on the various phenomena researchers call "collocation" to warrant using a specific existing definition. Section 2.5 will outline the assumptions of this study and present a new

definition of collocation, designed specifically for identifying learner collocations, to be used in the analysis of the data collected for this project.

2.1 Linguistic/lexicographic definitions of collocation

2.1.1 Lexical co-occurrence

At the heart of all definitions of collocation is the concept of co-occurrence; that is, the idea that certain words tend to occur together in text. This is the main assumption of collocation—that recurrence of word groups is a characteristic of language. The broadest definition of collocation, in fact, is nothing but this assumption. Sinclair holds to the following definition:

Collocation is the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening

. . . Collocation in its purest sense, as used in this book, recognizes only the lexical co-occurrence of words. (Sinclair 1991, 170)

For Sinclair, then, a collocation could consist of a pair of words such as *doctor* and *nurse*, just as long as they occurred together within a given span of text (which Sinclair limits to a maximum of four words on either side of the word in focus). As a linguist and lexicographer, Sinclair is interested in keeping his definition of collocation broad because he is looking for all possible patterns; to define collocation more narrowly is to lose important information.

There are advantages to a definition this broad: it can be adapted and used in quantitative studies, so that, for example, the frequency of co-occurrence of a pair of words can be measured against the frequency of each word individually in a corpus of texts. Many researchers involved in quantitative text analyses use this combination of word co-

occurrence and frequency as a springboard to other statistical measures. Delacourt (1992), for example, proposes an algorithm for computing co-occurrence statistics for a literary text. Smadja (1993a) has explored a large corpus of *The Wall Street Journal* with a computer program that generates degrees of word frequency which is then used to statistically define and predict word co-occurrence. Church et al. (1991) have devised a "mutual information statistic" which can be used to automatically identify commonly co-occurring sets of words. In the field of Natural Language Processing (NLP), quantifiable definitions are vital because of methodological limitations and the tremendous amounts of data required if one is to try to propose generalizations about the English language in its entirety. Work done by Chuang (1993), Church et al., Church and Hanks (1990), Clear (1994), Delacourt, Sinclair, and Smadja (1993b) all require quantifiable definitions of collocation.

Built into many definitions of collocation used in NLP applications such as automatic abstracting and indexing is the essential assumption that language is a stochastic process. That is, for the purposes of certain tasks, language is considered to be simply a set of randomly-occurring words. Though this assumption makes for convenient statistical conclusions about word co-occurrence, it disregards the element of context which language producers depend on in word selection. For the present study, however, this definition is too broad. The question addressed by this study involves more than simple co-occurrence.

Halliday, McIntosh and Strevens (1964) also understand collocation as the tendency of words to co-occur (33). In their view, grammar (formal patterns of language) fades gradually into lexis (a choice of many possible words), with collocation belonging on the latter end of the scale. Words sharing the same collocates are termed "lexical sets," which form patterns at the lexis end of the grammar-to-lexis spectrum. From this perspective, the phenomenon of collocation includes all words in a slot-filler model; unlike Sinclair, however, they view collocations as consecutively occurring groupings of words

of a language. Even so, collocations are not considered a subset of "set phrases," but as the probability each individual word in the English language has of being the filler of the next slot. Thus:

The formal criterion of collocation is taken as crucial because it is more objective, accurate and susceptible to observation than the contextual criterion of referential or conceptual similarity.

Lexical choice, therefore, is different from grammatical choice. In lexis, not only are there more items to choose from at any given point, compared with the items or classes one is choosing from in grammar; also there is no line to be drawn between those that can and those that cannot be chosen. (Halliday, McIntosh and Stevens 1964, 34)

Halliday's later work with Hasan (1976) develops the concept of collocation as simple co-occurrence of lexical items, emphasizing its cohesive role in language. However, as their work concerns lexical cohesion in whole texts, their focus is above the sentence level on word co-occurrences having a "lexicosemantic (word meaning) relation (285)." They describe their interpretation of the role of collocation as "any two lexical items having similar patterns of collocation—that is, tending to appear in similar contexts—[which] will generate a cohesive force if they occur in adjacent sentences (286)." In this view, collocation is not limited to pairs of words but in fact could appear as long chains of words interspersed among several sentences. The assumption of collocation as an inter-sentential element of language sets Halliday and Hasan's work apart from that of most others.

Related to the idea that certain word patterns are recurrent in important ways is the notion of arbitrariness: collocations are arbitrary co-occurrences (Fontenelle 1992:222, Smadja 1993a:146). While word co-occurrences can be calculated and predicted, not every set of statistically significant co-occurrences satisfy most definitions of collocation. So far, there has been no claim in the literature of a totally accurate search mechanism for

collocations based on syntactic structures or semantic representations, though Smadja has designed a computer program to identify collocations based partly on syntactic structures and partly on co-occurrence statistics. Additional evidence of the arbitrary nature of collocation lies in the fact that collocations are difficult for the native speaker to produce on demand. Studies by Berry-Rogghe (1973) and Seal (1981), however, show that native speaker subjects provided quite uniform answers on cloze exercises in which part of a collocation was omitted. In sum, though there are characterizations of collocations and their use, there are no known rules for generating them.

In an effort to further refine the concept of collocation, several researchers have differentiated between collocation and "significant" collocation, pointing to a gradient in which some collocations occur more frequently than others (Cowie 1978; Mackin 1978; and Seal 1981). Mackin, in his discussion of whether or not to include a given collocation in a learner dictionary, proposes that its "probability of co-occurrence" be considered (151-2). Seal defines significant collocations as "being those high-frequency collocations in which one of the items occurs with its syntagmatic partner in a particular construction at least ten per cent of the time (vii)." This distinction will be of no use in this study, because of the limited language domain in which this study takes place.

Aside from the basic assumption of co-occurrence, a definition of collocation usually involves one or both of the following considerations: grammatical relationships among the words in the collocation, and lexical (meaning) relationships among the words in the collocation. The importance of these two considerations is reflected in the fact that almost all definitions of collocation either put forward or used by linguists and lexicographers make reference to grammatical relationships or, more often, meaning relationships between the members of a collocation. Definitions taking grammatical relationships into account often divide a collocation into two distinct, unequal parts: a node and a collocate, where the node is specifically a "content" word (i.e., a noun, verb,

adjective, or adverb), and other words are the collocates. Definitions which take meaning relationships into account are more numerous, and tend not to divide the collocation into unequal parts.

2.1.2 Grammatical relationships

Interestingly, the definitions that rely solely on grammatical relationships tend to be much broader than those that rely on lexical features or both features. One example of a definition of collocation based on the assumption of co-occurrence and grammatical structure is given in Kjellmer (1987), "A collocation is a sequence of words that occurs more than once in identical form [in the Brown Corpus] and which is grammatically well-structured (133)." These conditions, that the words be in sequence, that they appear more than once, and that the sequence be grammatically well-structured, leave room for a broad range of results. Items such as *the house* or *a door* or *my car*, which occur in the Brown Corpus would have to be considered to be collocations under this definition. Definitions with implications similar to Kjellmer's include those used by Clear (1992; 1994) and Kennedy (1990). Seal (1981) uses a slightly narrower definition by limiting co-occurrence to especially high-frequency sets of items, and by not considering articles to be constituents of collocations. These are what Seal refers to as "significant collocations (vii)."

Benson, Benson, and Ilson divide collocations into two types, lexical and grammatical, based on a grammatical definition of collocations:

In English, as in other languages, there are many fixed, identifiable, non-idiomatic phrases and constructions. Such groups of words are called recurrent combinations, fixed combinations, or collocations. Collocations fall into two major groups: grammatical collocations and lexical collocations. (Benson, Benson, and Ilson 1986, xiv).

A grammatical collocation is a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause. (Benson, Benson, and Ilson 1986, vix)

Lexical collocations, in contrast to grammatical collocations, normally do not contain prepositions, infinitives, or clauses. Typical lexical collocations consist of nouns, adjectives, verbs, and adverbs (Benson, Benson, and Ilson 1986, xxiv)

Examples of lexical collocations would be the phrases, *work hard*, *heavy smoker* and *think highly*. Examples of grammatical collocations would include *give in to*, *make an effort*, and *by accident*.

This categorization provides for a much more precise definition which is especially useful for work in lexicography. In fact, Benson et al. have catalogued eight different types—or templates—of English grammatical collocations (with subtypes), and seven types of English lexical collocations. Though this scheme provides for a more constrained definition of collocation, there are so many collocations in general English (thousands of each type), that the resulting dictionary based on this definition is almost as large as a standard dictionary of single words.

2.1.3 Meaning relationships

2.1.3.1 Collocation as an element of meaning

The issue of a lexical or meaning relationship among the components of a collocation is viewed from different perspectives by various researchers. Firth, often given the credit for bringing the word "collocation" into common linguistic parlance, envisions collocation as an aspect of meaning. He contends that part of the meaning of a word arises from the words it collocates with:

I propose to bring forward as a technical term, meaning by 'collocation', and to apply the test of 'collocability'. The following sentences show that part of the meaning of the word *ass* in modern colloquial English can be by collocation:

- (i) An ass like Bagson might easily do that.
- (ii) He is an ass.
- (iii) You silly ass!
- (iv) Don't be an ass! (Firth 1957, 194-195)

Firth's point is that in each of the above examples the word *ass* has a slightly different shade of meaning. Regardless of the theoretical viability of the claim that meaning is derived in part via collocation, the idea that collocation is connected to meaning is not wholly implausible. It is in fact supported by many of those who have studied collocation in different contexts, among them Newmark (1973), a translator who proposes that collocation involves the restriction of at least one of the words in the set by eliminating other possible meanings. He explains, for example, that *pay attention* is a collocation because both *pay* and *attention* are restricted in meaning (1973:9). Unlike Newmark, Schenk, whose work centers around the computer-assisted translation of collocations and idioms, asserts that only one of the words retains its "central" meaning (the base) while the other word (the collocate) gains its meaning from context (1994:4-5).

Researchers assuming different theoretical frameworks of language also have attempted to account for an element of meaning in collocation. Bloksma and van der Kraan propose a computational approach to "implementing ideas on collocation" (1992:21) within a Categorical Grammar framework, where the head word (base) and the collocate are analyzed differently. They claim that a collocate "has a meaning that is different from its 'neutral' sense or it is semantically empty (20)." Müller has also studied the semantic aspect of collocations. His (1981) stratificational analysis of collocation yields a binary distinction

between collocations whose members are attracted by a set of semantic restrictions, and collocations whose members are drawn together by force of "conventional usage (178)." In the former category Müller supplies *the girl with tousled hair* and *the man passed away*; examples of the latter category are *last night* and *a strong likelihood* (178). In other words, Müller claims that *tousled hair* and *passed away* are collocations because they follow language-specific semantic restrictions, while *last night* and *a strong likelihood* are collocations not because of semantics, but because of random choices native speakers have made to use these combinations more frequently. It is this last group of collocations, according to Müller, that NNSs have the greatest problems producing.

2.1.3.2 Collocations as lexical units

There are several more issues regarding meaning relationships in collocations. One of these is concerned with the concept of "lexical unity." Are collocations lexical units? That is, do the members of a collocation cohere more closely than members of any random phrase? Empirical evidence suggesting an affirmative answer exists in the form of a study by Harris (1995), who investigates the "unit status" of collocations by testing the word recognition skills of native speakers. Her findings indicate that native speakers detect collocations with the same facility with which they detect single words, and that the detection of random pairs of words takes longer than the detection of collocations. Collocational dictionaries give tacit assent to these conclusions (see, e.g., Sinclair 1987b and Benson et al. 1986), by giving word groupings a place of importance equal to that of individual words.

Lexical unithood is expressly mentioned in several linguistic and lexicographic studies of collocation. Benson et al. (1986) define collocation in part by asserting that the native speaker will reject violations of lexical unity. Benson (1985) includes lexical unity in his definition of grammatical collocations and compares them in that regard to idioms,

which he also regards as lexical units. Smadja (1993a) gives "cohesive lexical clusters" as one of his four characteristics of collocation (147). The assumption that collocations are stored as wholes in the minds of native speakers of a language is held by descriptive linguists Bolinger (1975) and Müller (1981). Lexical unity is an important concept to collocation because it expresses the idea that words in these phrases cannot be substituted for. So, for example, murder must be *committed*, not *performed* or *done*. There is a cohesive element, as Smadja puts it, to these phrases—a cohesive element that is purely semantic. Müller would agree with Smadja about the fact that *commit* and *murder* form a collocation, but he would disagree that the underlying tie was semantic. For Müller, *commit murder* is a collocation due to syntagmatic restrictions.

Two similar explanations for lexical unity constraints—the difficulty with replacement—have appeared in the literature as semantic constraints. Newmark (1973), writing about collocation in the context of a translation problem, contends that "...collocates within a collocation define and delimit each other by eliminating at least some of their other possible meanings; the defining may be mutual and equally balanced, but more often it is closer for one collocate than for another (9)." Cruse (1986), a semanticist, seconds this notion, pointing out that collocations "...do have a kind of semantic cohesion—the constituent elements are, to varying degrees, mutually selective (40)." Semantic constraints are the cause of the "lexical unity" described above; lexical unity is the concept that no other word will do, but "semantic cohesion" is the reason why. Compare, for example, the collocation *reasonable price* with the phrase *convenient price*. Not only is *reasonable* the only word that will fit; only a certain sense of the word *reasonable* will fit. A replacement word either lacks the particular semantic shading required by the collocate, and/or the replacement does not evoke the same shading required by the original member. This word cannot be replaced—because each of the words in the collocation very specifically limits the others in terms of semantic shading.

2.1.3.3 Collocations as meaning analyzable structures

Numerous references have been made to the difference (or lack thereof) between collocations and idioms. Benson (1985), Bolinger (1975), Cruse (1986), Fontenelle (1994), Schenk (1994) and Smadja (1993a) all adhere to a strict distinction, based at least on the differences in the derivation of their meanings. Collocations are held to be transparent in the sense that the meaning of the whole phrase can be derived from the meanings of the individual words, whereas idioms are opaque and must be understood metaphorically. For example, the words *put down* in the collocation *put down the book* and the idiom *put down the rebellion* have a meaning recoverable from context in the first instance but not in the second. Sinclair (1991), while agreeing in principle to this distinction, maintains that the line between collocation and idiom is fuzzy.

Herbst (1996:388-89) strongly objects to transparency as the distinctive feature of collocation on two counts. First, he thinks that some collocations can be analyzed as either transparent or opaque, because there is no one objective method of semantic analysis. Therefore, we cannot classify all collocations as meaning analyzable structures. In other words, he is saying that there are collocations (e.g., *commit murder* and *false teeth*) whose elements are not motivated to appear together on the basis of meaning. Secondly, Herbst believes that there are collocations in which the two words of a pair share such a close semantic affinity that transparency is not an issue; these words have such a small subset of collocates that analysis of meaning is not necessary. Herbst supplies the example *flock of birds*, where *flock* has very few nouns with which it can combine.

Some, like Mitchell (1971) and Müller (1981) claim that at least part of the difference between collocation and idiom lies in the degree of "productivity." In this view, idioms are seen as frozen, unable to change in terms of transitivity, substitution,

expansion, transposition, and the like. Collocations, then, are co-occurrences with some variability allowed. However, while it is true that many idioms are rigid structures, Cowan (1989) has no trouble supplying exceptions to this criterion: *hit/strike the high point*, *keep/hold up one's end*, and *shoot the bull/breeze*, for example. At best, then, the issue of productivity is only a distinguishing feature in some cases.

Some other researchers (most notably Nattinger and DiCarrico 1992 and Lewis 1993) have put together strong cases against a binary distinction between collocation and idioms and instead hold that the relationship between the two is that of a continuum. According to Nattinger and DiCarrico, idioms are an "ossified" form of collocation, which means that they are predictably frozen forms which have lost meaning because of their invariability. The more a form allows for different types of variation, according to Nattinger and DiCarrico, the more meaning it gains until at the far end of the continuum it is a "free combination" (i.e., just a pair of words that happen to occur together in text). This view presents collocation as a life-cycle: a collocation begins as a free combination and as it becomes more commonly used it becomes more fixed until it becomes fixed enough to be termed an idiom.

Gitsaki's (1996) extensive study on NNS acquisition of collocations includes a comprehensive review of published studies on the differences between collocations and idioms. Her conclusion is that semantic transparency is the only criterion that could be reasonably used as a distinguishing factor. In Gitsaki's study, a distinction between the two is not important; in this study, however, the distinction between collocations and idioms is quite significant because they fulfill different roles in language. The idiom is used to describe, in a few words, a situation which has some analogical connection to the idiom; it is an association of meaning between a given situation and that from which the idiom was derived. The collocation, by contrast, is a conventionalized way of expressing a certain idea; it has to do with semantic shades of words and their conventionalized uses. Farghal

and Obiedat (1995) contend that "...other things being equal, common collocations...are more useful to L2 learners than idioms... (319)." A second language learner could get along quite well without producing idioms (though he or she would have to learn idioms in order to understand them), but they would be misunderstood or ineffective in relaying information, especially in a restricted language domain, if they did not produce the correct collocations. Since there is a distinction between the use of collocation and idiom, and a distinction in their usefulness to the second language learner, there should be a clear-cut distinction made for the purposes of this study. The most useful distinction, and the one lent most credence by researchers, is that of transparency.

2.1.4 Examination of assumptions

An observation of previous work on the concept of collocation reveals three commonly-held assumptions: 1) that collocation is a linear phenomenon (i.e., there is a certain order to the parts of a collocation), and/or the examination of collocation should proceed in linear fashion; 2) that the concepts "frequency" and "familiarity" are both necessary conditions for collocation; and 3) that native-speaker data, intuition, and categorization is the only perspective from which to view collocation.

Many linguistic theories view lexical choice as an item-by-item operation. Halliday, McIntosh and Strevens (1964), for example, in their contention that collocation is a statistical phenomenon, assume that words are all just fillers of slots, and that they are chosen by the language producer one-by-one. This is an assumption needed for item-by-item text analysis in the field of natural language processing. It is opposed to Sinclair (1991) and those who follow him. For Sinclair, there are two mechanisms for the construction of language: what he calls the "open-choice principle," in which a single word is selected; and what he believes is the far more prevalent mechanism of language production, the "idiom principle," in which the language producer in a single choice selects

a "chunk" of language. When analyzed in this way, the errors of the NNS make sense—they simply do not know that a particular set of words forms a unit in the English language, or they use the open-choice principle where the idiom principle is necessary.

The analysis of collocation is also often seen as a linear process. Automated searching for collocation is done by comparing the number of co-occurrences of words on either side of a word under scrutiny. The larger the data set, the more accurate the results are assumed to be.

The second frequently-held assumption has to do with a mixing of the concepts of "frequency" (i.e., an absolute count of how often a word appears in a text) and "familiarity" (i.e., patterns in text recognized by the native speaker). For example, the definition of collocation espoused by many computational linguists, that collocation is a set of words that commonly co-occur, is undergirded by the belief that frequency bespeaks familiarity. In other words, because a set of words or a phrase is familiar to the native speaker (i.e., seen as a collocation), that set of words or phrase occurs frequently in text. Are there collocations that are familiar but not so frequent? Could either or both of these be necessary but not sufficient conditions for collocation? In any case, familiarity is the more meaningful condition, because assumptions about frequency run directly counter to the experience of the non-native writer, who may see a collocation many times, but is unable to correctly reproduce the phrase from memory. So this assumption, while perhaps valid for the native speaker, may not hold true for learners of English, a fact which must be considered in a definition of collocation for this study.

Native-speaker bias appears in the data used in collocation studies and in the assumptions themselves that are built into these studies. For example, many descriptions/analyses of collocation involve no empirical data, but simply the writer's native-speaker intuition (see, e.g., McIntosh 1961; Mitchell 1971; Müller 1981; Herbst 1996). Native-speaker intuition certainly provides insight into various aspects of

collocation, but impedes insights from other corners at the same time. Native-speaker-produced texts are used almost exclusively in empirical analyses of collocation, and because collocation is difficult to spot, identification of collocation is reduced to frequency criteria. If native-speaker intuition is to be used it should be used at its point of strength, which is to correctly derive a collocation when presented with just a portion of it. The way native speaker intuition would be put to best use in this study is by using it to spot problems in collocations produced by non-native speakers.

In addition to the work done by researchers in the fields of lexicography and linguistics, it is worth considering studies and definitions of collocation in the ESL/ESP literature, which draw from the above studies, but are more likely cover the specific types of issues I will encounter in my exploration of collocation. What follows is first a defense for examining collocation in specific language domains, and then a review of relevant ESL/ESP literature.

2.2 Studying specific domains

Most studies of collocation have focused on what the ESL teacher would refer to as "general English"—the entire spectrum of language domains that make up the English language. Little work has been done on specific language domains and collocations. In his studies of genre, Swales (1990) proposes 6 characteristics of discourse communities, one of which is shared, specialized lexis. He does not mention collocations by name, but does imply collocation belongs in this category. In separate research reports, linguist Kjellmer (1987) and lexicographer Smadja (1993b) claim that collocations are especially characteristic of restricted language domains. Moreover, according to Knowles (1993), collocations are of vital importance to advanced learners of English who are seeking to become part of the conversation in restricted language domains. The graduate student in business, for example, needs to be capable of producing the collocations specific to the

graduate business English domain in order to gain membership in that discourse community.

Focusing on a specific domain, in this case business English, is important for two reasons. First, it is a much easier and more reasonable task. Second, the larger the domain, the larger the study needed to test it and the more variables there are to take into account. In addition, more general studies are difficult to replicate.

2.3 English for specific purposes

Before laying out the assumptions and definition of collocation to be used in this study, it is necessary to consider what the ESL/ESP literature says about collocation.

2.3.1 References with an overt definition of collocation

Almost all definitions of collocation written by and for ESL/ESP professionals, though they may draw on work by linguists, lack the theoretical depth and detail offered by researchers in other fields. Interestingly, the most insightful definition in ESL/ESP literature is one of the oldest. H.E. Palmer (1933) saw collocations through the grid of NNS errors. His own experiences with Japanese learners of English led him to a phenomenon that he termed "collocation" and which he described as "successions of words that must or should be learnt, or are best learnt as integral wholes rather than pieced together from their component parts (10)." Unfortunately, Palmer's work has rarely been cited and therefore his insights into collocations have been largely missed by ESL professionals.

Most definitions that do appear in the ESL/ESP literature lack not only Palmer's grasp of the nature of the learner's difficulty, but also the broader complexity of collocations. *The Grammar Book* by Celce-Murcia and Larsen-Freeman (1983:54-55), for

example, is a standard reference grammar of English for ESL teachers, but refers to collocation as simply "...those words that frequently co-occur in a predictable syntagmatic relationship." Though Johns is a well-respected researcher in business English, her study of written business discourse (1980:39) follows Halliday and Hasan in viewing collocation as a "catch-all term for all lexical features not coded elsewhere." Long's article on the potential of teaching collocation as part of vocabulary describes collocation this way:

To say that words collocate means nothing more than that certain words repeatedly and predictably occur alongside other words. In other words, any pair or group of words occurring or fitting together in this predictable way is a collocation. Collocate and collocation can be extremely useful words in the classroom, if taught with the warning that they are classroom words, and have no validity outside the classroom. (Long 1990, 37)

There are several books specifically directed at the teaching of vocabulary which address the issue of collocation. Taylor (1992) devotes an entire chapter to collocation, for which she provides a broad definition. For Taylor, words go together "either because of meaning similarities or because they stand beside each other in an accepted order combination (77)." In other words, collocation encompasses both semantic sets of words and recurrent word combinations, including idioms. Though she does not discuss the problems the NNS faces in collocation, she does provide several classroom activities that are intended to help the NNS explore the connections between words.

Nation (1990) also includes a section on collocation in his *Teaching and Learning Vocabulary*. While his material is classroom-activity directed, he does provide a narrower definition of collocation than Taylor. His suggestions for classroom activities are based on what has been reported in the literature: grids and gap-filling.

One small volume targeting NNS themselves is *American Business Vocabulary* (Flower and Martínez, 1995). The book consists of a series of thematic exercises designed

to be worked through independently by NNSs, and includes useful items that distinguish, for example, between *make* and *do*. Unfortunately, there is little in the way of explanation about collocation. However, these may well be quite useful in the hands of the teacher who already understands the importance of collocation.

Hatch and Brown (1995) discuss collocation in *Vocabulary, Semantics, and Language Education*. They see collocation as part of the set of lexical phrases that make up the "phrasal lexicon." As a result, they do not offer a detailed look at collocation, except as it relates to other phrasal units such as reduplication, compounds, idioms, and proverbs. The only mention of collocation in regard to the NNS involves a discussion of how synonyms have different collocates. They do offer exercises, though they are not specifically geared to the NNS.

Much clearer and theoretically-based definitions of collocation for the vocabulary teacher are found in Carter (1987) and McCarthy (1990). Carter in particular gives a very thorough overview of existing definitions of collocation, as well as descriptions of how collocation can affect style and grammar. Carter's suggestions for teaching collocations include both grids and gap-filling exercises; he does not, however, explore the nature of NNS errors. While Carter's theoretical treatment of collocation is comprehensive, McCarthy's may be more accessible to the language teacher. McCarthy sees collocation as the central aspect of vocabulary study, and provides specific examples of NNS difficulty with them. Carter also suggests gap-filling and semantic domain exercises.

ESL/ESP journal articles containing definitions that do recognize the complexities involved in collocation tend to dismiss them as unimportant, or simplify them to a large degree. Zughoul (1991:51) concludes in a study of learner writing that collocational errors are "directly related to transfer from the native language." Dechert and Lennon (1987) admit to a perplexing divide between lexical and grammatical considerations in the analysis of

NNS collocational errors, but fail to enumerate them and in fact conclude that transfer must somehow be the root of the problem.

Korosadowicz-Struzynska (1980) is one of the few ESL professionals to thoroughly cover a definition of collocation in a journal article. She proposes three criteria: 1) that they are set expressions whose meanings can be deduced by the learner from context or background knowledge; 2) that they are less closed to alteration than are idioms; and 3) that they do not constitute idioms in the learner's own language. The problem with this definition is that it circumscribes a different set of phrases for each different native-speaker group. This kind of definition may work well in the EFL classroom, where all students normally have the same native language, but would prove unwieldy for the teacher in an ESL context, where students have multiple native languages.

2.3.2 References with no overt definition of collocation

Several ESL/ESP authors provide no overt definition of collocation. In Bowen et al. (1985), for example, no explicit definition is provided, though collocation as a language phenomenon relevant to second language learning is discussed at length. In an article about teaching vocabulary, specifically collocations and their possible replacements, Martin (1984) also fails to define collocation. Presumably, the authors of such pieces are unaware of the variability in meaning the term carries.

In light of the fact that some researchers in the disciplines mentioned above have constructed definitions to suit particular areas of inquiry, while others have either avoided defining collocation or have defined it in a simplistic manner, there is no single definition that this study will build on. Instead, a clear and reasoned definition (vital to the credibility of this project), will be presented as a set of five characteristics.

2.4 Assumptions of this study

In light of the above observations, collocation as used in this study does not rely upon a previously proposed specific definition. Instead, a set of characteristics taken from those definitions is proposed to help identify NNS collocations. For the purposes of this study collocation is defined as two or more co-occurring words within a sentence which:

1. Are in a syntagmatic relationship (i.e., are not simple co-occurrences such as *doctor - nurse*).
2. Manifest a non-grammatical semantic constraint in at least one of the members (e.g., *near* in *near future* is limited to a time sense).
3. Together embody a meaning recoverable from the component parts (i.e., is not an idiom).
4. Do not allow for replacement items; valid replacement constitutes a separate collocation, even if meaning remains constant (for example, *reasonable price* and *affordable price* will be considered to be separate collocations).
5. Are seen by the native speaker as a lexical unit (i.e., the native speaker instinctively knows when one of the above constraints has been violated).

Frequency of co-occurrence, an important criterion in some definitions, will be considered to be irrelevant here since the domain of English is limited to business English in an MBA context, and the corpus of data is limited. What can be claimed, then, is that the results gained through the analysis of this data are representative of this individual situation, but the implications will suggest possibilities of further studies in other specialized domains.

This chapter has reviewed prevailing definitions of collocation and their underlying assumptions. It has been shown that these definitions are not sufficient for a study of the type undertaken here; instead, a definition of collocation strictly for the purpose of identifying learner collocations (i.e., those collocations produced by NNSs that are seen by

native speakers as idiosyncratic) has been suggested. The chapter 3 will explain the general background of the study and how the above definition was used in the identification of learner collocations.

CHAPTER 3

DATA AND METHODOLOGY

3.0 Introduction

This chapter specifies as well as justifies the methodology of the present study. It describes the data collection context and procedures, the coding of data, and the analytical tools and their role in proposing an answer to the central question motivating this study: How can learner collocations (i.e., idiosyncratic collocations produced by NNSs) in business texts written by native speakers of Mandarin Chinese be characterized and accounted for?

This chapter is divided into seven sections. Section 3.1 provides a general background of the use of corpora in research, with special attention to choices made in the creation of the corpus examined herein. Section 3.2 explains the context in which the data were collected. Section 3.3 describes the participants involved in the study. Section 3.4 outlines the data types collected during the study: a student background questionnaire, student writing, and student reading. Section 3.5 explains the data collection procedure. Section 3.6 describes how collocations were gathered from the corpus and stored for further analysis, including an explanation of the computer tools used for these purposes. Section 3.7 spells out the coding rationale and procedures. Chapter 3 ends with a short conclusion.

3.1 The use of corpora in research

3.1.1 Corpus creation and limitation

Engwall states that there are two ways to create a corpus: the chance method and the choice method (1994). The chance method involves selecting examples as they are encountered in one's reading—gathering isolated texts which appear to be good data. The problem here is that these samples may or may not be representative of a text type. The choice method involves finding a subpopulation of texts which accurately represents the given text type. According to Engwall, depending upon what sort of generalizations are to be drawn, researchers must carefully define a domain and criteria for corpus sampling if the results of a corpus study claim to hold generalizations for the given text type.

Engwall holds to the choice method, and suggests a procedure for the selection of texts to be included in a corpus. The first choice should be made in terms of category (e.g., literary works, newspapers, and so forth); the second step is the choice of genre or subgenre (e.g., drama, prose, scientific texts, and so forth); the third choice is one of period, or time frame (e.g., synchronic vs. diachronic); the final step is to decide on the actual texts to be included in the corpus or, if the texts are too long, then the actual samples of those texts (i.e., given choices made for each of the three categories above, what texts or text fragments match these criteria?).

There is of course a tension in representativeness between the breadth (how many different texts) and the depth of the corpus (how much of each text is included). In other words, given a limited corpus size a researcher must decide which is more important: the breadth of coverage afforded by including only fragments from a large number of different texts, or the depth provided by a smaller number of whole texts. Ultimately, according to Engwall, these choices must be based on the objectives of the research. de Haan agrees, saying, "the suitability of the sample depends on the specific study that is undertaken, and that there is no such thing as the best, or optimum, sample size as such (1992:3)."

Engwall's selection procedure was used in choosing the texts which comprise the corpus in this study. The text selection procedure is explained below.

The biggest problem in studying NNS texts written in a language sub-domain such as engineering or biology is availability: tracking down NNS texts authored under similar conditions is quite difficult. Few genre-specific English classes are taught in universities, with the exception of business English. Business English classes for NNSs are often offered in universities with graduate business programs. The availability of written texts in the domain of business English was the deciding factor in the determination of which context to use to observe learner collocations. According to Engwall's text selection procedure, then, the category choice for this study is NNS written texts; the logical genre choice is business English.

In terms of time frame, the length of the university semester (15 weeks) dictated that this study be primarily synchronic. Engwall's final two choice points, text and sample, conflate in this study, since the texts were collected and used in their entireties. In addition to the choices explained above, the fact that this particular study included students with different native languages required that a further choice in corpus creation be made.

According to Granger (1993), a pioneer in NNS corpus research, the literature on learner English is often based on insufficient data and differences in research methods, which has led to contradictory findings. She therefore advocates drawing data from a learner corpus in which a large amount of standardized data can be analyzed by many different researchers using similar methods, which may lead to more reliable generalizations about learner errors. The thrust of Granger's corpus research is to separate errors based on L1 transfer from errors that are not L1-dependent. To accomplish this, she makes an effort to compile a corpus that is balanced in terms of different L1 learner compositions. Granger's concern about L1 influence is taken up in this paper in terms of choices made about the composition of the corpus. Because it was not possible to balance a

corpus for the present study in the way Granger recommends, different L1 influences on the corpus were eliminated by limiting the texts included in the corpus to those produced by NNSs who shared the same native language.

Therefore, in addition to corpus creation choices suggested by Engwall, a decision was made after the corpus of texts was collected to limit the analyzed portion of the corpus to just those texts authored by native speakers of Mandarin Chinese.

3.2 Data-gathering context

The data for this project were produced during the Fall 1995 semester in the context of a special English class for 13 International students: The English Language Institute at UT Arlington's MBA Bridge Program. Twelve of the participants in this study had been accepted by the Graduate School of Business, with the provision that they obtain the English writing skills necessary to negotiate a graduate business curriculum. The thirteenth student was enrolled in the Graduate School of Engineering and decided to participate in the MBA Bridge Program of his own volition. All 13 students were highly motivated to improve their writing skills, since continued enrollment in their major fields of study demanded that they improve their English writing skills.

The MBA Bridge Program at UT Arlington was initiated in the Spring 1995 semester in response to an expressed need by graduate business students for an English program separate from UT Arlington's Intensive English Program. They felt, and administrators agreed, that a specialized business English curriculum of reading and writing would best help them attain the English proficiency required for successful completion of graduate business work. The business English writing class met 2 hours each day, Monday through Friday, for the entire semester. The curriculum included grammar structure practice, writing practice, and critical reading and writing skills. All of the assignments

were in some way related to types of assignments students would face in a graduate business classroom.

The instructor of the MBA Bridge class, while aware of the fact that written texts were being collected for examination and analysis, was not informed of the nature of this study. This was done to insure that he did not "accidentally" divulge information to students, or change his normal teaching or marking practices in any way that would influence the outcome of this study. After all data had been collected, the students were informed about the exact nature of the study, the types of data that were being examined, and the possible outcomes of the study.

In order to gather, examine, characterize and account for learner collocations in the writing of the participants in the study, several types of data were collected: a language and academic background questionnaire, a corpus of student writings, and a corpus of student reading assignments on which the writing assignments were often based. All of these data types are explained in detail below.

3.3 Participants

While 13 students were enrolled in the Bridge Program, only those texts produced by the eight native speakers of Mandarin Chinese were used in the present study. The corpus was narrowed to include only texts produced by native speakers of Mandarin for three reasons. First, they comprised the largest language group in the class and therefore the results of any analysis of the class writings would be heavily influenced by them. Second, it seemed a logical move after an examination of preliminary data showed that the Chinese speakers were producing similar learner collocations. Third, the researcher has studied Chinese and is familiar with Chinese structures, allowing for a more thoroughgoing analysis of collocations produced by the Chinese speakers.

Students in the Bridge Program were placed there based upon their performance on a 30-minute composition test, referred to as the entry exam. The instructor of the Bridge Program course graded the compositions on a scale of 1-6+; those students receiving a grade of five or below were required to enroll in the class. Any student who completed the exit exam at the end of semester with a score of six or higher was deemed to have erased their English deficiency and was no longer required to be enrolled in the course. The entry exams were not collected as part of the corpus, but the exit exams were. Table 1 lists the eight study participants by their code initials with their corresponding entry and exit composition scores.

TABLE 1
STUDENT ENTRY AND EXIT SCORES

Student Name	Entry Score	Exit Score
CA	5	6
CH	5	6
DI	5	6
GO	4	5
LU	4	6
LY	5	6
PO	4	5
WI	5	6+

A preliminary outline of this study, as well as a request for participation by students in the study, was approved by The Human Subjects Committee of UT Arlington (see

appendix A). All students involved in the study signed an informed consent form at the beginning of the Fall 1995 semester.

3.4 Data types

Three types of data were collected in the course of this study: a questionnaire completed by each of the participants, written texts produced by each of the participants, and a set of texts each participant was required to read. Each of these data types is explained in detail below.

3.4.1 The questionnaire

The first type of data was collected from a written English questionnaire designed to elicit information about the students' backgrounds (see appendix B). The questionnaire was comprised of 21 questions, which requested information such as age, sex, native language, exposure to English and other languages, TOEFL and GMAT scores, undergraduate major, current class load, and the amount of time spent reading outside of class. Table 2 displays information from several of the questions for each student.

TABLE 2
STUDENT BACKGROUND INFORMATION

Name	Gender	Age	TOEFL	GMAT	English Study	Country	Other Languages
CA	Male	26	550	520	7 years	Taiwan	No
WI	Female	25	553	540	13 years	Taiwan	No
LY	Female	27	540	550	10 years	Taiwan	No
GO	Male	25	510	520	8 years	Taiwan	No
LU	Male	26	540	560	10+ years	Taiwan	No
DI	Female	22	560	550	10 years	Taiwan	Japanese
PO	Female	28	510	490	10 years	Taiwan	Taiwanese Japanese
CH	Male	23	550	560	8 years	Taiwan	No
Mean	---	25.25	539	536	9.5 years	---	---

The study population is comprised of four males and four females, all born and raised in Taiwan. The range in age is six years, with 25.25 years as the average. TOEFL (Test of English as a Foreign Language) scores, designed to gauge a student's preparedness for study in an English language environment, range from 510 (two students) to 560 (one student). The average score for the study population is 539. A score of 550 is a standard minimum requirement for entrance into the graduate school at UT Arlington. The GMAT (Graduate Management Admissions Test), required of all applicants to the business school, is "designed to help graduate schools assess the qualifications of applicants for advanced study in business and management (ETS 9)." The scores attained by participants in this study range from 490 (one student) to 560 (two students) and the average is 536. The business school at UT Arlington does not have a published minimum acceptable score

on the GMAT; instead, they gauge admission status by weighing the GMAT score together with undergraduate transcripts. The students in this group all had scores high enough to gain conditional admission to the business school. The average number of years of formal English study is 9.5, with one student reporting as few as seven years of study and one student reporting as many as 13. Only two students reported having studied other languages: in both cases the language was Japanese. One of these two also reported being a speaker of Taiwanese.

3.4.2 Student writing

A copy of every assignment written during the semester, both in-class and outside assignments, was collected. There were a total of 19 writing assignments, including two drafts of a lengthy research paper. Due to absenteeism and time constraints, not all students turned in all assignments.

Five of the nineteen writing assignments are rewrites of original compositions. While this is a potential problem in terms of a duplication of data, and therefore duplication of learner collocations, these particular assignments are different enough that their inclusion does not constitute simply rewriting data to the corpus. This variable was accounted for in the analysis of the data by reducing learner collocations into types data. In other words, analyses were performed twice: once with all learner collocations and once with each different learner collocation counted only once.

Each assignment is outlined in table 3; the readings referred to in table 3 are listed in section 3.4.3. Immediately following table 3 is table 4, which shows each student's text production (in terms of number of words) for each assignment.

TABLE 3
WRITING TASKS

Task	Date	Description of the Assignment
1	09/07/95	This assignment was a summary of a reading given as homework the previous day (Reading #1).
2	09/11/95	This assignment was an in-class re-write of Task #1.
3	09/15/95	A summary written in class based on an article read that same day in class (Reading #2).
4	09/25/95	A short problem-solution paper written in class after instruction in how to write a problem-solution paper (Reading #4).
5	10/17/95	A take-home assignment which was a revision and expansion of the problem-solution paper written for Task #4.
6	10/18/95	This assignment was a take-home summary of Reading #5.
7	10/23/95	Second revision of the problem-solution paper from Task #4.
8	10/26/95	This assignment was a take-home critique of Reading # 6.
9	10/29/95	This assignment was a take-home summary of Reading #7.
10	11/01/95	This assignment was a take-home summary of Reading #8.
11	11/08/95	An individual summary of an article written outside of class as part of the process of preparing to write a research paper.
12	11/10/95	A second individual summary of an article written outside of class as part of the process of preparing to write a research paper.
13	11/13/95	An individual critique of an article written outside of class as part of the process of preparing to write a research paper.
14	11/15/95	A second individual critique written outside of class as part of the process of preparing to write a research paper.
15	11/18/95	Third revision of the problem-solution paper from Task #4 .
16	11/22/95	First draft of individual research paper.
17	12/06/95	Final Essay, done outside of class in response to Reading #10.
18	12/07/95	Exit essay question written in class in response to Reading #9.
19	12/08/95	Second Draft of Research Paper.

TABLE 4
NUMBER OF WORDS PER WRITING TASK

Task	CA	WI	LY	GO	LU	DI	PO	CH	Total
1	90	---	---	---	170	65	61	90	476
2	116	---	---	---	158	---	73	124	471
3	144	152	86	115	---	87	106	111	801
4	274	320	334	281	164	284	184	291	2,132
5	671	502	535	485	464	413	776	761	4,607
6	208	220	190	383	420	258	275	259	2,213
7	1098	763	749	595	1005	570	997	1228	7,005
8	362	280	343	---	132	147	188	202	1,654
9	237	126	304	---	128	119	300	124	1,338
10	---	164	221	120	175	97	203	161	1,141
11	424	126	254	189	74	62	214	248	1,591
12	191	115	236	127	187	89	165	312	1,422
13	174	184	259	---	168	---	---	324	1,109
14	230	188	302	112	145	---	228	262	1,467
15	1446	900	876	550	1116	660	1161	1292	8,001
16	2939	2824	2148	1864	1707	1831	2420	1563	17,296
17	478	458	294	279	457	202	354	467	2,989
18	229	163	195	198	188	180	186	226	1,565
19	3378	3798	2485	1684	2269	2187	2687	2481	20,969
Total	12,689	11,283	9,811	6,982	9,127	7,251	10,578	10,526	78,247

There is a wide range in the total number of words produced by each student over the course of the semester. GO produced the fewest number of words, with 6,982. At the other extreme is CA, who produced a total of 12,689 words. However, for all students approximately half of total production was generated for the two drafts of the research paper. The total number of words in the corpus is 78,247.

3.4.3 Student reading

Most writing assignments were cued to a reading assignment; the reading was designed to be completed before the writing assignment was attempted. Though it has been hypothesized that NNSs have difficulty producing collocations in part because they lack exposure to them in reading material (Hussein 1990, Long 1990, Zughoul 1991), that factor could not be accounted for here, because it is beyond the scope of this study. The readings, then, constitute an environmental factor that was not controlled. Table 5 shows the date of each assignment, the title, and the number of words in the text. In addition to the listed readings, each student read a variety of academic research materials in preparation for writing a research paper.

TABLE 5
CLASS READING ASSIGNMENTS

Reading No.	Date	Title	Word Count
1	09/06/95	Reading on Marketing	552
2	09/15/95	How to Really Help People Escape Poverty	481
3	09/21/95	Market Targeting	1,084
4	09/25/95	A Problem Solving Case: Marketing	169
5	10/03/95	Consumer Behavior: Basic Concepts	3,856
6	10/24/95	A Welcome Sign for Immigrants	897
7	10/25/95	Preparing for the Worst	1439
8	10/31/95	China Turns to Confucius for Help with Ethical Crisis	830
9	12/04/95	Job Skills or Knowledge?	468
10	12/05/95	We Should Cherish our Children's Freedom to Think	776
11	12/07/95	Don't Call Me, I'll Call You—but Not From Cloud 9	732
12	12/07/95	The Fear of Dying	1819

3.5 Data collection

Papers were copied immediately after the class period in which they were turned in. They were then rendered into an electronic format by the researcher, who either keyed them in (if handwritten) or scanned them (if typewritten). In both cases, papers were carefully

checked for spelling errors, which were corrected if the intended word was discernible. Each assignment for each student was saved to a separate file for later processing.

3.6 Data processing: Gathering and storing the collocations

This section will discuss the three methods of analyzing the corpus of learner collocations: reading through the corpus texts (section 3.6.1), using a concordancing program (section 3.6.2), and using the full text database program FolioVIEWS (section 3.6.3). In addition, section 3.6.4 will discuss the storage and analysis of collocations using FileMaker Pro, a relational database program.

3.6.1 Gathering collocations through reading

The first attempt at gathering collocations from the texts was handled at the paper-and-pen level; thus the first view of collocation—the traditional view, or "flat" view—was that gained through simply reading text produced by the NNS. This step entailed reading straight through each of the individual texts, noting collocational irregularities, and highlighting them for later inclusion in a database. Irregular collocations are often readily identified because they jump out at the native speaker, who immediately recognizes that there is something wrong with phrases such as **tax rate basket* and **advocate his ground*. Unfortunately, this is an inefficient and potentially inaccurate method and it quickly became apparent that an automated method of gathering collocations would be necessary.

3.6.2 Gathering collocations with a concordance

The greatest problem with gathering learner collocations by simply reading through texts is that it is possible to overlook more subtle collocational errors. It is also possible for any researcher who is reading through texts to simply pass over a collocational error due to

the monotony of the task. A concordance, on the other hand, provides a more focused view from which to identify irregular collocations, because it allows the researcher to see irregularities in groups.

Before the advent of computers, a concordance was generally available for certain well-studied books, such as the Bible. Now, however, there are quite a few freely available concordancing programs which allow the user to produce a concordance for any set of electronic texts. The program used in this study to create a concordance of all student writings is called "Conc." Conc provides the means to focus, in alphabetical order, on each of the content words present in a document or set of documents. Note that in figure 1 it is easy to see that the student had trouble using the word *potential*.

line	pre-word span	focus word	post-word span
99	price. This means if	Polio	wants to promote a new cologne
114	analyze these data,	Polio	has to make some decision for its
120	winner or looser. Even	Polio	now faces the difficulty in men's
41	we know, one of the	Polio's	basic products is aimed at
45	are the largest	population	who really buy the products
60	revealed the potential	possibility	of developing student's
27	to find out the most	potential	buyers, Polio conducted one
59	result revealed the	potential	possibility of developing
57	respondents, 36 of them	prefer	the cologne. As we can see from
28	Polio conducted one	preliminary	marketing survey according
50	the result of the	preliminary	survey, Polio decided to

Figure 1. Conc Concordance Fragment.

For this study, each student's complete set of texts was folded into one file, and a concordance was made of each student's set of writings. The concordance was then used to examine spans of text (a 'span' being four words to the left and right of the word under scrutiny) for collocations not seen during the flat level of analysis. This arrangement of the texts focuses on each word in each student's set of texts one at a time, ensuring that nothing gets missed. Not only does the concordance view of the student texts reveal more

subtle collocational errors, it also places all the spans containing the same words together, making word usage patterns and problems for each student readily visible. Only content words were focused on at this level; irregular collocations were cut and pasted directly into word processing documents to protect against any errors in copying.

3.6.3 Gathering collocations with FolioVIEWS

The computer program FolioVIEWS, a free-text database environment, was used to assist in the display and analysis of three types of data collected in this study: student-produced texts, assigned readings, and student questionnaires. A general overview of the merits of FolioVIEWS is given here; an explanation of how FolioVIEWS is used to help organize and code the data types used in this study is given below.

The aspects of FolioVIEWS which make it especially amenable to this project are the visual mark-up, query and record-keeping features. Text characteristics including highlight colors and font types can be assigned to labels which can in turn be applied to any section of text that fits user-defined criteria. Sections of the corpus can be grouped so that, for example, all texts by one author, or written on one date, can be viewed or queried separately; the infobase software indexes each word of the entire corpus individually, making any word, root, synonym, or antonym easy to find. This record-keeping feature is perhaps the most valuable for text analysis, allowing the researcher to return to the data in context whenever necessary, with just a few keystrokes. A backtracking option allows the user to review each action taken in the current session. These tools combined with the researcher's intuition and imagination provide the opportunity to explore the corpus in a manner not possible in a pen-and-paper context, as well as to create a visual map of the results of the analysis.

3.6.3.1 Student-produced texts

Any study of learner collocations requires that the broad context of such phrases be available for review, because such a context is most often the key to determining the collocation a native speaker would use. In this study, the NS collocation was required for two separate procedures: coding learner collocations involving semantic variation and eliciting back translations, which were used to determine cases of direct translation. FolioVIEWS creates from the student texts a flexible database in which all original text files retain their integrity, and can be viewed as a linear set of texts, but also one that continually allows for a perspective of the collocations in their broader context.

FolioVIEWS was used to mark up the student texts to make the learner collocations apparent at a glance. After texts were placed in the database, the collocations identified using pen-and-paper and the concordance were marked with color and font variation, so that upon visual inspection the irregular collocations could easily be seen. Learner collocations were marked differently depending on the probable cause(s) of the idiosyncrasy: phonology, grammar, translation, or semantic variation. In the coding stage, known learner collocations were placed in a FolioVIEWS query to determine whether other students had written the same phrase. This cross-checking of idiosyncrasies over the entire corpus sometimes brought to light previously unrecorded collocations. Other learner collocations were revealed when the texts were examined with a series of queries designed to detect synonyms and roots of elements of collocations already identified.

There is no single method to use when searching for learner collocations. In other words, the researcher must follow leads provided by results at previous levels, as well as her own intuition, in deciding how to proceed. In this case, for example, the researcher had previously observed learner collocations involving expressions about the concept of quality (e.g., **increase the quality of life*). A synonym query on the word *quality* brought to light the irregular **rate of consumers*, which in its context means *number of consumers*. So

while *rate of consumers* was passed over unnoticed at both the traditional (flat) and concordance levels of analysis, at the query level *rate* was identified on a conceptual basis (though perhaps accidentally), and the irregular word grouping detected.

Another query type which displays the power of a free-text database analysis is the word root query. Based on the researcher's belief that competition, compete, and words with similar roots were significant enough in the student writings to warrant an examination, a query based on the root *competi-* was run and produced 137 hits. Of those, two were irregular collocations not identified at other levels of analysis: **high competition (high-level competition)* and **gain more competition (become more competitive)*.

Finding collocational variation with FolioVIEWS is not purely a matter of intuition, however. This third step is also a check to make sure any collocations identified at a previous level are checked for all authors and for any suspicious elements across the entire corpus. For example, three irregular collocations by a single author were identified at the first and second levels involving the word *advantage*. A simple query to the infobase revealed 45 hits for this word in the entire corpus, and a survey of those 45 turned up a new learner collocation, **receives more advantage* for a different author.

Once irregular collocations are marked in the database they can be grouped, viewed, and annotated in any number of ways. For example, hypertext links can be made between like collocations, or between collocations which express similar concepts. Thus, the database can be queried for all learner collocations involving the concept of time, and then each of the resulting hits can be hypertextually linked to the others, allowing the comparison between different contexts to be permanently available. In addition, expandable windows containing the user's notes can be created next to each paragraph; in this case, the researcher used those windows to hold the presumed correct versions of learner collocations appearing in the corresponding paragraph.

From the traditional pen-and-paper view, the researcher's perspective is limited to one place in one text at one time. The computerized concordance perspective allows the researcher a more thorough analysis, including an inspection of how each individual author used each word in an entire text. It also simplifies the process of identifying collocations that an author revised at some point during data collection. The free-text database affords additional analytical tools: all authors' works are simultaneously laid open with query functions that identify patterns in word root and word sense use. More importantly, the visual markup features of the database create a visual map of the corpus, which is invaluable in helping the researcher identify items of interest at a single glance.

3.6.3.2 Student-assigned reading texts

The second type of data stored in the FolioVIEWS database is the set of student readings. These were organized by date in the order in which they were read by the class. This set of texts was queried for correct forms of collocations for which students had idiosyncratic forms. They were also hypertextually linked to both erroneous and correct forms in the student-produced text corpus.

3.6.3.3 Student questionnaires

The third type of data stored in the FolioVIEWS database is the set of student questionnaires. These data were used to correlate learner collocation types with questionnaire variables such as age, TOEFL and GMAT scores, number of years of English study, and so forth.

3.6.4 Using FileMaker Pro to store and examine collocations

Though FolioVIEWS holds all learner collocations in their original contexts, there are advantages to viewing the isolated collocations, and to sorting them according to features such as author, type of idiosyncrasy, and so forth. Every phrase identified in the FolioVIEWS database as an idiosyncratic collocation was placed in a Filemaker Pro database. Each collocation has its own record, for which is listed the author, true collocation (NS collocation), original source (from the first attempt at coding according to Lombard 1995), any notes on grammar, phonology or translation, and semantic categories. An example is displayed in figure 2.

Diss.file		
Collocation Entries		
Entry Copy	Author Chuan-Ming Lu	Source
	True Collocation the internal company structure	<input type="radio"/> Missing Member
Records: 571	Error the interior of company structure	<input type="radio"/> Blends
Found: 2	Notes	<input type="radio"/> Word Order/Gram
Unsorted	Phonology final syllable	<input checked="" type="radio"/> Semantic Misapp
	Translation	
	Semantics typical adjective	
	<input checked="" type="checkbox"/> typical adjective	<input type="checkbox"/> LabReal
	<input type="checkbox"/> standard praise for entry	<input type="checkbox"/> Liquify
	<input type="checkbox"/> continue	<input type="checkbox"/> Magnification
	<input type="checkbox"/> culmination of	<input type="checkbox"/> Manifestation
	<input type="checkbox"/> degradation of	<input type="checkbox"/> Minus
	<input type="checkbox"/> excessive funct	<input type="checkbox"/> Multiple
	<input type="checkbox"/> Fact "realization of" verb	<input type="checkbox"/> harm/injure/impair
	<input type="checkbox"/> stopping of	<input type="checkbox"/> Obstruct
	<input type="checkbox"/> Function	<input type="checkbox"/> Operator
	<input type="checkbox"/> Hypernym/hyponym	<input type="checkbox"/> Permit/allow
	<input type="checkbox"/> Core of	<input type="checkbox"/> Plus
	<input type="checkbox"/> verb (non-participant involvement)	<input type="checkbox"/> Positive attributes
	<input type="checkbox"/> Inception	<input type="checkbox"/> Prep. for in/from/to
	<input type="checkbox"/> Labor	<input type="checkbox"/> Copula for nouns/adjectives
		<input type="checkbox"/> Preposition
		<input type="checkbox"/> Proximity
		<input type="checkbox"/> Highly prob:
		<input type="checkbox"/> Realize
		<input type="checkbox"/> Typical nou
		<input type="checkbox"/> Typical instr
		<input type="checkbox"/> Typical loca
		<input type="checkbox"/> Typical mea
		<input type="checkbox"/> Typical moc
		<input type="checkbox"/> Typical resu
		<input type="checkbox"/> Singular
		<input type="checkbox"/> Verity

Figure 2. Collocation Database Record.

3.7 Coding

3.7.1 Sources of collocational idiosyncrasy

Once learner collocations had been gathered, several categorization schemes were considered, including the researcher's own taxonomy of irregular collocations (Lombard 1995). In fact, this 4-element taxonomy of collocations was used in the initial coding procedure. Learner collocations were assigned to a "surface source": missing member, blends, word order/grammar error, and semantic misappropriation. These categories, however, were not revealing enough of the nature of the learner collocations. Therefore, they were broken down into type by (probable) cause. Four general sources of idiosyncrasy became apparent: idiosyncrasy due to lexical transfer (from L1); idiosyncrasy involving phonology; idiosyncrasy involving non-standard grammar; and idiosyncrasy involving semantic variation. Since it was impossible to determine what each student was thinking when they produced each learner collocation, all possible causes were considered for each irregular collocation. Each category is explained below.

3.7.1.1 Idiosyncrasy due to lexical transfer

Learner collocations were placed in this category when they matched back translations provided by a native speaker of Mandarin. The back-translations, plus the researcher's own knowledge of Chinese, enabled the identification of content words that were translated directly from Mandarin into the English collocation. For example, **open the computer*, **a long time trip*, and **borrow urgent funds*, were all identified through back translations as cases of direct lexical transfer from Mandarin Chinese. All learner collocations were checked against back translations before they were considered to be candidates for any other category, because it has been hypothesized that native language transfer is a major cause of learner-type collocations.

3.7.1.2 Idiosyncrasy involving phonological similarities

Collocations were placed in this category because of a dissimilarity in sound between a word contained in the learner collocation and a corresponding sound in the target collocation. These idiosyncrasies were observed on the phone level (where either a vowel sound or a consonant sound difference was noted) as well as on the syllable level (entire syllable differences occurred in syllable-initial, syllable-medial, and syllable-final positions). For example, **distinctive grope of people*, **tax rate basket*, and **to reserve this drawback*, were all identified as learner collocations involving phonological idiosyncrasies. In all cases idiosyncrasies in phonology resulted in a phonologically similar lexical item.

3.7.1.3 Idiosyncrasy involving non-standard grammar

From another vantage point these learner collocations might be diagnosed as grammar errors only, having nothing to do with collocation. The cases of learner collocation involving irregular grammatical structure examined in this study reflect a restricted semantic sense which closely approximates what is required for the context. In other words, students appear to know the meaning of the target collocation, but not the structure. For example, **care price*, **a trend of change*, and **consuming behavior* are close approximations of native speaker collocations, with the exception of a structural variation (i.e., a missing preposition, an unusual modification, and an irregular word form). In these cases, the student has composed a phrase conveying the same general meaning as the corresponding NS collocation but has failed to package it in a structure which, to the researcher, is native-sounding.

3.7.1.4 Idiosyncrasy involving semantics

Semantic idiosyncrasy occurs when a collocation is constructed correctly in grammatical terms, retains a slightly different semantic shading than the context calls for, but is not a case of direct lexical transfer from the native language. This category reflects students' closest approximations to native speaker collocations. For example, **products are pointed*, **value concepts*, and **impending era*, were identified as learner collocations because semantically they are only slightly different from their native speaker equivalents (*products are aimed*, *value judgements*, and *coming age*, respectively). The students producing these items know the meaning and structure of the collocation, but may not apprehend the semantic shading of one of the members. Of the 571 learner collocations identified in this study, 69% were due to semantic idiosyncrasy.

3.7.2 The categorization of semantic idiosyncrasy

The overwhelming number of learner collocations diagnosed as semantic variation dictated the use of an established categorization scheme that would provide for a more finely-grained analysis. Of the different semantic approaches to the lexicon, only a relational approach to the lexicon makes explicit a structural organization that other semantic models take as implicit. Within the relational approach exists the concept of lexical relation: A lexical relation is a culturally recognized pattern of association that exists between lexical units in a language (LinguaLinks 1997). Lexical relations specify how a given word combines with other words to express certain relationships, such as magnitude (a tall man, versus a high man*) and a decrease (reduce taxation, versus save taxation*). These are the types of relations (and combinations thereof) that will be used to classify the various learner collocations found in the student-produced texts.

Relational semantic perspectives have been used to solve many problems in lexicography (see Evens 1988). For example, the new LinguaLinks program designed and

produced by the Summer Institute of Linguistics for field linguists working on bible translation includes dictionary tools which allow the program user to select from among a general set of lexical relations in constructing dictionary entries (LinguaLinks 1997).

Of the many researchers exploring relational semantics in lexicography, Mel'čuk (1986) is known for his pioneering dictionary work. Mel'čuk and his colleagues, intent on designing a dictionary that would supply language learners with all the information they would need to write like native speakers, includes in each entry of his Explanatory Combinational Dictionary a section called "lexical functions" (i.e., lexical relations). According to Mel'čuk and Zholkovsky (1988), the lexical functions section of the dictionary "[characterizes] the idiomatic, that is, language-specific substitutability and co-occurrence relations of the entry lexeme...(55)." Based on their work in Russian, Mel'čuk and Zholkovsky posit a set of standard basic lexical functions, which they claim can be adapted for use with any language.

Because of the success other researchers have had with Mel'čuk's lexical functions (see chapter 2), they were adopted for use in this study. Each of the semantically-based learner collocations found in the data were classified according to the Mel'čuk lexical function expressing the relation present in the collocation. Of Mel'čuk's 53 categories, 20 were used in this study. Table 6 lists each of the functions used in this study in the left-hand column, and in the right-hand column lists a matching explanation of its relational property with an example in the form $f(x) = y$. Note that these categories are not necessarily mutually exclusive.

TABLE 6

LEXICAL RELATIONS USED IN THIS STUDY (from Frawley 1988)

Function	Example
A ₁ , A ₂	typical adjective for numbered participant A ₁ (suspicion) = full of
Bon	standard praise for entry Bon (advice) = sound
Culm	culmination of Culm (ability) = peak
Excess	excessive functioning of Excess (eyelid) = flutter
Func	verb which takes the entry as subject of the first participant Func (idea) = come to
Incep	the beginning of Incep (fly) = take off
Labor	verb which takes entry as secondary object Labor (esteem) = hold (x holds y in esteem)
Liqu	the elimination of Liqu (group) = disband
Magn	intensity (quantity or duration) Magn (hatred) = deep
Manif	is manifest in, entry as subject Manif (tear) = well up
Minus	less of Minus (wind) = slacken
Mult	a regular aggregate of Mult (paper) = ream
Obstr	to function with difficulty Obstr (justice) = obstruct
Oper	verb which takes entry as object Oper (party) = throw
Plus	more of Plus (joy) = grow
Prox	to be on the verge of Prox (disaster) = on the brink of
Real	verb meaning to realize with entry as object Real (ambition) = realize

Once all the collocations of semantic origin were labeled in the FilemakerPro database, the analysis could begin.

Initially, each of the learner collocations was placed into one of the four collocation types (lexical transfer, phonological, grammatical, semantic). Later, each entry in the

Filemaker database was revised to include all possible categories under which the given idiosyncratic collocation could be placed. So, for example, the entry for **interior of company structure*, originally coded as involving phonological idiosyncrasy (*interior* replacing the intended *internal*), was revised to contain a semantic coding as well. Once the database revision had been completed for every entry, the Filemaker database could be searched for the number of hits involving a pair of collocation types (e.g., all collocations coded for both semantic and phonological influence could be retrieved). This revision of the database was necessary for determining frequencies required by the analysis described in chapter 5.

3.8 Conclusion

This chapter has established the validity of the data collection context and procedures, focusing especially on the following points:

1. How corpus data were gathered, and how these methods can be used in the present study.
2. How various computer applications were used in data collection and coding.
3. How Mel'čuk's set of lexical relations (lexical functions) can be used to label learner collocations involving semantic variation.

Chapter 4 presents the results of the analysis of the corpus for learner collocations.

CHAPTER 4

RESULTS AND ANALYSIS

4.0 Introduction

This chapter describes the results of the collocation analysis of the corpus. The presentation of the results is organized around each of the four types of collocational idiosyncrasy found in the student corpus: lexical transfer of a word or phrase, phonological similarity, grammatical irregularity, and semantic choice. In sections 4.1 through 4.4, a general description of each learner collocation type is provided, a listing of all idiosyncratic collocations found in the corpus is given, and at the end of each section a summary of the number of idiosyncratic collocations for that learner collocation type is reported for each student and writing task.

This chapter is structured so that the learner collocation types are discussed in order from least to most productive. As explained in chapter 3, all idiosyncratic collocations were checked first through back translation for evidence of lexical transfer: if a phrase was not a candidate for the lexical transfer category, it was then considered for the phonology category, and then for the grammar category. All collocations that were not strong candidates for these three categories were then considered for the semantic category. Each collocation was placed in the category deemed most likely to represent the primary reason for the observed variation.

While each idiosyncratic collocation collected in this study has been placed in a single variation category, there are many cases in which the idiosyncrasy could involve the influence of two or more categories. This aspect of the study will be discussed in chapter 5.

In addition to the four abovementioned categories, there is a "questionable" category of collocations, left to be examined later. The questionable cases are noted at the end of section 4.5.

4.1 Collocational variation involving primarily lexical transfer

Lexical transfer is a significant factor in the Mandarin speaker's construction of English collocations. In this study, lexical transfer refers to the reliance on the native language in selecting lexical items which become members of English language collocations. This strategy is evidenced by the student's use of words that, directly translated, would be appropriate in the speaker's native language.

Idiosyncratic collocations involving lexical transfer discovered in this study actually comprise two separate types of lexical transfer from Mandarin to English. The first type is the direct word-for-word translation. In other words, the subjects of this study often made substitutions of directly translated Mandarin words or phrases into their written English, causing idiosyncratic collocation (e.g., *open the computer, *long time trip, *accelerating old). Table 7 lists all idiosyncratic collocations of this type with the code for the student who wrote each one (ST in each of the tables in chapter 4). Following the student code is the number of times a particular phrase occurred in parentheses. Bold-faced type represents actual points of lexical transfer.

TABLE 7

IDIOSYNCRATIC COLLOCATION INVOLVING DIRECT TRANSLATION

ST	Student Phrase	Target Collocation
CH	most potential environment disaster	greatest potential ... disaster
LY	most potential buyer	greatest source of potential buyers
LY	the most potential buyers	buyers with the greatest potential
LU (3)	the most potential market	the largest potential market
LY	the most potential buyers	the most likely buyers
LY	the most possible buyers	largest group of potential buyers
PO	open the computer	turn on the computer
GO	open the monitor and computer	turn on the monitor and computer
GO (2)	a certain time range	a certain time frame
LU	a long time trip	a time-consuming trip
CH	accelerating old	aging rapidly
CH	will improve with getting old	will improve with age
CH	extending life expectation	increasing life expectancy
CH (2)	who are above 55	who are older than 55
LY	forever winner or looser	complete/final winner or loser
CH	arrange workers	schedule workers
DI (2)	fewer-children marriages	smaller families
GO	make a conclusion	draw a conclusion
PO (3)	make the good influence on the world	have a good influence on the world
LY	severe competition	intense/stiff competition
LU (2)	the female aspect	the female angle/segment
CA (2)	borrow urgent funds	borrow emergency funds

The second type of lexical transfer involves two English words with distinctly separate meanings and uses which are expressed in Mandarin by a single word. So, for example, the words *make* and *do* caused numerous idiosyncratic collocations because in many contexts Mandarin has only one word, *zuo*, representing what *make* and *do* express in English (though note that in table 7, the two instances of *make* are direct lexical borrowings; they do not involve the word match problem discussed here). Six of the eight students in the study produced a *make/do* error in at least one of their papers.

A significant number of the word match category collocations involve numbers. Number-related collocations cause problems for at least two reasons. First, Mandarin has no special word for *once* or *twice*; one possible variation arising from Mandarin to English lexical transfer are the idiosyncratic collocations **twice times* and **once at a time*.

Secondly, Mandarin expresses fractions differently than English, naming the denominator first, then the numerator, and always using ordinal numbers. For example, the Mandarin rendering of *one-third* would be *san fen zhi yi* (literally, three measure of one). English, on the other hand, uses cardinal numbers in the numerator and ordinal numbers in the denominator, producing numbers such as one-third, two-fifths, and so forth. In this study, students sometimes wrote out fractions in collocations as phrases like, **third-fourth*, **four-five*, and **three-four*.

In tables 8 and 9, all idiosyncratic collocations of the word match type are listed along with the code for the student (ST) who wrote each one, and the number of times a particular phrase occurred in parentheses. Bold-faced type represents the lack of a match between Mandarin and English.

TABLE 8

IDIOSYNCRATIC COLLOCATION INVOLVING WORD MATCHING

ST	Student Phrase	Target Collocation
DI (2)	a survey made by	a survey done by
DI (2)	the survey made by	the survey done by
DI (3)	a survey was made	a survey was done/performed
GO	experiments had previously been made	experiments had ... been done
GO (2)	make a market survey	do (take) a market survey
CA	make a marketing survey	do a marketing survey
CA	make marketing survey	do a marketing survey
CA	make some of the analysis	do some of the analysis
CA	making another survey	doing another survey
LY	we made an investigation	we did/conducted an investigation
LY	scientists make a research	scientists do research
LU	the survey should be remade	the survey should be redone
PO	efforts which have been done	efforts which have been made
CA (2)	change in strategy should be done	change in strategy should be made
CH (2)	do some change	make some change
LY	the dollars will not always keep down	the dollar will not always stay down
LY	US dollars might keep weak	US dollar might stay/remain weak
LU	making us keep in a small area	making us stay in a small area
LY (2)	do not in favor of	do not favor + are not in favor of
LY	are tend to in favor of	tend to favor + are in favor of

TABLE 9

IDIOSYNCRATIC COLLOCATION INVOLVING NUMBER MATCHING

ST	Student Phrase	Target Collocation
WI	add... once at a time	add...one at a time
CA	multiply marketshare twice times	multiply marketshare twice/two times
WI	nearly twice times more	nearly twice as much/two times more
LY	around twice times of students	around twice as many students
LY	around twice times of them	around twice as many of them
CA (2)	over third-fourth of students	over three-fourths
LY	over four-five of the students	over four-fifths of the students
LY	three-four of students	three-fourths of students
LY	three-four of consumers	three-fourths of consumers

A summary of the number and type of idiosyncratic collocations involving primarily lexical transfer for each student is shown in table 10.

TABLE 10
 IDIOSYNCRATIC COLLOCATION INVOLVING PRIMARILY LEXICAL
 TRANSFER

ST	Direct Translation Variation	Other Lexical Transfer Idiosyncrasy		Total
		Word Match	Number Match	
CA	2	6	3	11
WI	0	0	2	2
LY	6	7	5	18
GO	4	3	0	7
LU	6	2	0	8
DI	2	7	0	9
PO	4	1	0	5
CH	7	2	0	9
Total	31	28	10	69

Table 11 contains a summary of the idiosyncratic collocations involving lexical transfer recorded in the corpus, according to the number produced by each student in each writing task, all of which are listed and explained in chapter 3. Each task is designated "in" for work written in class, or "out" for work written outside of class. Dashed lines in a cell represent papers that were not handed in. Rows in bold text represent an assignment which elicited ten or more idiosyncratic collocations involving primarily lexical transfer. A total of 69 idiosyncratic collocations involving primarily lexical transfer were recorded over the entire student-produced corpus.

TABLE 11
 LEXICAL TRANSFER COLLOCATION SUMMARY

Task	CA	WI	LY	GO	LU	DI	PO	CH	Total
1 in	0	--	--	--	0	0	0	0	0
2 in	0	--	--	--	0	--	0	0	0
3 in	0	0	0	0	--	0	0	0	0
4 in	0	0	0	0	0	0	0	0	0
5 out	1	0	4	2	2	1	1	0	11
6 out	0	0	0	0	0	0	0	0	0
7 out	4	1	4	1	2	1	1	0	14
8 out	0	0	0	--	0	0	0	0	0
9 out	0	0	1	--	0	0	1	1	3
10 out	--	0	0	0	0	0	0	0	0
11 out	0	0	0	0	0	0	0	0	0
12 out	1	0	0	0	0	0	0	0	1
13 out	0	0	0	--	0	--	--	0	0
14 out	0	0	1	0	0	--	0	0	1
15 out	3	0	5	0	2	1	1	0	12
16 out	1	0	2	2	0	3	0	5	13
17 out	0	0	0	0	0	0	0	0	0
18 in	0	0	0	1	2	0	1	0	4
19 out	1	1	1	1	0	3	0	3	10
Total	11	2	18	7	8	9	5	9	69

Table 12 displays the number of idiosyncratic collocations per one hundred words for each assignment. Each task is designated "in" for work written in class, or "out" for work written outside of class. Dashed lines in a cell represent papers that were not handed in. Cells with 0.5 or more idiosyncratic collocations involving primarily lexical transfer per 100 words are indicated in bold.

TABLE 12
 LEXICAL TRANSFER COLLOCATIONS PER 100 WORDS

Task	CA	WI	LY	GO	LU	DI	PO	CH	Ave
1 in	0	--	--	--	0	0	0	0	0
2 in	0	--	--	--	0	--	0	0	0
3 in	0	0	0	0	--	0	0	0	0
4 in	0	0	0	0	0	0	0	0	0
5 out	0.14	0	0.75	0.41	0.43	0.24	0.13	0	0.24
6 out	0	0	0	0	0	0	0	0	0
7 out	0.36	0.13	0.53	0.17	0.20	0.18	0.10	0	0.20
8 out	0	0	0	--	0	0	0	0	0
9 out	0	0	0.33	--	0	0	0.33	0.81	0.22
10 out	--	0	0	0	0	0	0	0	0
11 out	0	0	0	0	0	0	0	0	0
12 out	0.52	0	0	0	0	0	0	0	0.07
13 out	0	0	0	--	0	--	--	0	0
14 out	0	0	0.33	0	0	--	0	0	0.07
15 out	0.21	0	0.57	0	0.18	0.15	0.09	0	0.15
16 out	0.03	0	0.09	0.11	0	0.16	0	0.32	0.07
17 out	0	0	0	0	0	0	0	0	0
18 in	0	0	0	0.51	1.10	0	0.54	0	0.26
19 out	0.03	0.03	0.04	0.06	0	0.14	0	0.12	0.05
Total	0.09	0.02	0.18	0.10	0.09	0.12	0.05	0.09	0.09

4.2 Collocational idiosyncrasy involving primarily phonological similarity

Phonological similarity was a more productive source of error than lexical transfer. In this study, phonological similarity refers to confusion encountered by the student at the phone and syllable level in selecting appropriate words found in English language collocations. This strategy is evidenced by the student's use of words that are only slightly different in sound than what would be appropriate in the given context. Though these idiosyncrasies have been judged to involve primarily phonological idiosyncrasy, there may be other forces that led to the phrases listed in this section. The issue of combined motivations will be discussed in chapter 5.

These phonologically-based idiosyncrasies occur at two different levels: the phone level, where the difference between the idiosyncrasy and the target word is a single sound, and the syllable level, where the difference between the idiosyncrasy and the target word is an entire syllable. For example, one student wrote, **expend the market share*, instead of *expand the market share*; another wrote, **a lower tax rate basket* instead of *a lower tax bracket*. In addition, there were a small number of occurrences where a one- or two-syllable word contained multiple points of phonological idiosyncrasy. These were classified as word-level idiosyncrasies.

Phone-level idiosyncrasies involve either the substitution, deletion, or addition of certain vowel or consonant sounds to the target word. Most of the idiosyncrasies result in real words. A total of twenty-two phone-level idiosyncrasies were found in idiosyncratic collocations in the corpus: twelve involving vowel sounds, and ten involving consonant sounds. Table 13 shows all phone-level idiosyncrasies found in the corpus listed along with the code for the student (ST) who wrote each one, and the number of times a particular phrase occurred in parentheses. Letters in bold-faced type represent the phone

that is substituted or the addition of a sound. Phrases without bold letters signify that a sound was dropped; that sound is rendered in bold in the corresponding target phrase.

TABLE 13

PHONE-LEVEL IDIOSYNCRATIC COLLOCATION

(a) Vowel sound idiosyncrasy

ST	Student Phrase	Target Collocation
CA	bo nd to think	bound to think
LY	market is fill of uncertain factors	market is full of uncertain factors:
LY	market is fill of variable risk	market is full of variable risk
CA	expend the market share	expand the market share
LY	safe he av en in a sea of global trouble	safe haven in a sea of global trouble
CH	to ad opt ourselves to	to adapt ourselves to
LU (2)	ad opt ed past experiences	adapted past experiences
CA	the price of food will r is e	the price of food will rise
GO	r is e moral standards	raise moral standards
LY	r is e the interest rate	raise the interest rate
CH	distinctive gr o up of pe o ple	distinctive group of people

(b) Consonant sound idiosyncrasy

ST	Student Phrase	Target Collocation
CA	put in a b l ind	put in a bind
CH	target marketing is w i ldly used	target marketing is widely used
WI	releas e the po v erty problem	relieve the poverty problem
WI	installed over t i mes	installed over time
LY	two-third	two-thirds
WI (3)	three-fourth students'	three-fourths of students'
CA (2)	four fifth of all ... loans	four-fifths of all. .. loans

idiosyncrasies which resulted in entire syllables either substituted, deleted, or added to the target word of a collocation were deemed to be syllable-level idiosyncrasies. These substitutions, additions, or deletions appeared in three different parts of multi-syllable words: five appeared in the initial syllable, nine appeared in the medial syllable, and twenty-seven occurred in the construction of the final syllable. In table 14 is a list of all syllable-level variations found in the corpus listed along with the code for the student (ST) who wrote each one, and the number of times a particular phrase occurred in parentheses.

Letters in bold-faced type represent the syllable that is substituted or added; phrases without bold letters signify that a sound was dropped, and that syllable is rendered in bold in the corresponding target phrase.

TABLE 14

SYLLABLE-LEVEL IDIOSYNCRATIC COLLOCATION

(a) Initial syllable idiosyncrasy

ST	Student Phrase	Target Collocation
CA	in order to make references	in order to make inferences
PO (2)	a lower tax rate basket	a lower tax bracket
CH	revive the irreversible damages	survive the irreversible damages
GO	wastes of sources	waste of resources

(b) Medial syllable idiosyncrasy

ST	Student Phrase	Target Collocation
LY	dollars are relevantly weaker	dollars are relatively weaker
LY	dollars are relevantly weak	dollars are relatively weak
CA	keep the serving of the loan	continue servicing the loan
CA	provide the loan serving	provide loan servicing
PO (3)	graduation commitment	graduation commencement
WI	adjusting themselves to survive	adapting themselves to survive
PO	the reflection from competitor	the reaction from competitors

(c) Final syllable idiosyncrasy

ST	Student Phrase	Target Collocation
CA (2)	relatively investing concepts	relevant investing concepts
LY	relative with people's life	relevant to people's lives
LY	dollars are relevant weak	dollars are relatively weak
LY (2)	dollars relevant weak	dollars are relatively weak
LY	decline in relevant to	decline in relation to
CA	to reserve this drawback	to reverse this drawback
CH (2)	distinct themselves from others	distinguish themselves from others
LU	to distinct those who use this product from others	to distinguish those who use this product from others
CH (2)	the main strain of the society	the mainstream society
CH (2)	mainstreams of the economy	mainstays of the economy
LY	to adjust itself	to adapt itself
DI	adjust themselves	adapt themselves
LU (2)	the interior of company structure	the internal company structure
LU (2)	the products are various	the products are varied
LY	differentiate percentage	different percentages
PO	how to result the problem	how to resolve the problem
CH (2)	moral conscious	moral conscience
DI	purchase and use luxurious items	purchase and use luxury items
LY	intensive competition	intense competition

The final set of phonological idiosyncrasies were labeled word-level idiosyncrasy, because they either involved the entire word or contained multiple points of sound substitution/addition. These idiosyncrasies are shown in table 15, which lists all instances of word-level idiosyncrasy in idiosyncratic collocations found in the corpus, listed along with the code for the student (ST) who wrote each one, and the number of times a particular phrase occurred in parentheses. Words in bold-faced type represent a word that is phonologically similar to the target.

TABLE 15

WORD-LEVEL IDIOSYNCRATIC COLLOCATION

ST	Student Phrase	Target Collocation
CA (2)	failure to rescue the loss	failure to recoup the loss
CA (2)	Polio will rescue the loss	Polio will recoup the loss
CH	their physical fail	their physical frailty
CH	physical failing	physical frailty
CA (2)	lower loan-to-value rations	lower loan-to-value ratios
GO	a new specific product	a specialized product

A summary of the number and type of phonological similarity idiosyncrasy for each student is shown in table 16.

TABLE 16

IDIOSYNCRATIC COLLOCATION INVOLVING PRIMARILY PHONOLOGY

ST	Phone Level		Syllable Level		Similarity	Word	Total
	Consonant	Vowel	Initial	Medial			
CA	3	3	1	2	3	6	18
WI	5	0	0	1	0	0	6
LY	1	4	0	2	8	0	15
GO	0	1	1	0	0	1	3
LU	0	2	0	0	5	0	7
DI	0	0	0	0	2	0	2
PO	0	0	2	4	1	0	7
CH	1	2	1	0	8	2	14
Total	10	12	5	9	27	9	72

Table 17 contains a summary of the phonological similarity idiosyncrasy recorded in the corpus, according to the number of variations made by each student in each writing task. Rows in bold text represent an assignment which elicited ten or more idiosyncratic collocations involving phonological similarity. A total of 94 idiosyncrasies due to phonological similarity were recorded over the entire student-produced corpus.

Following the learner collocation summary table is table 18, displaying the number of idiosyncrasies per one hundred words for each assignment. Cells with 0.5 or more collocations per 100 are indicated in bold.

TABLE 17
 PHONOLOGICAL SIMILARITY COLLOCATION SUMMARY

Task	CA	WI	LY	GO	LU	DI	PO	CH	Total
1 in	0	--	--	--	0	0	0	0	0
2 in	0	--	--	--	0	--	0	2	2
3 in	1	1	0	0	--	0	0	0	2
4 in	1	0	1	0	0	0	0	0	2
5 out	1	1	1	1	0	0	2	1	7
6 out	0	0	0	0	1	0	0	0	1
7 out	2	1	0	0	1	0	1	1	6
8 out	0	0	0	--	0	0	0	0	0
9 out	0	1	0	--	0	1	1	1	4
10 out	--	0	0	0	0	0	0	0	0
11 out	0	0	0	1	0	0	0	0	1
12 out	1	0	0	0	0	0	0	0	1
13 out	0	0	0	--	0	--	--	2	2
14 out	0	0	2	0	0	--	0	1	3
15 out	2	1	2	0	1	0	1	2	9
16 out	7	0	6	0	2	0	1	2	18
17 out	0	0	0	1	0	0	0	0	1
18 in	0	0	1	0	0	0	0	0	1
19 out	3	1	2	0	2	1	1	2	12
Total	18	6	15	3	7	2	7	14	72

TABLE 18

PHONOLOGICAL SIMILARITY COLLOCATIONS PER 100 WORDS

Task	CA	WI	LY	GO	LU	DI	PO	CH	Ave
1 in	0	--	--	--	0	0	0	0	0
2 in	0	--	--	--	0	--	0	1.61	0.42
3 in	0.69	0.66	0	0	--	0	0	0	0.12
4 in	0.36	0	0.30	0	0	0	0	0	0.09
5 out	0.15	0.40	0.19	0.21	0	0	0.26	0.13	0.09
6 out	0	0	0	0	0.24	0	0	0	0.05
7 out	0.18	0.13	0	0	0.10	0	0.10	0.08	0.09
8 out	0	0	0	--	0	0	0	0	0
9 out	0	0.79	0	--	0	0.84	0.33	0.81	0.30
10 out	--	0	0	0	0	0	0	0	0
11 out	0	0	0	0.53	0	0	0	0	0.06
12 out	0.52	0	0	0	0	0	0	0	0.07
13 out	0	0	0	--	0	--	--	0.62	0.18
14 out	0	0	0.66	0	0	--	0	0.38	0.20
15 out	0.14	0.33	0.68	0	0.09	0	0.09	0.15	0.11
16 out	0.24	0	0.28	0	0.12	0	0.04	0.13	0.10
17 out	0	0	0	0.36	0	0	0	0	0.03
18 in	0	0	0.51	0	0	0	0	0	0.06
19 out	0.08	0.03	0.08	0	0.09	0.05	0.04	0.08	0.06
Total	0.14	0.05	0.15	0.04	0.08	0.03	0.06	0.13	0.09

4.3 Collocation idiosyncrasy involving primarily irregularities in grammar

Idiosyncratic collocations involving problems with grammatical structure are significantly more numerous than those involving either native language transfer or phonological idiosyncrasy. For the purposes of this study, grammatical variation is defined as variation in the structure of a given collocation. In other words, students used word orders, word forms, prepositions, genitive structures, and so forth that do not conform to the rules of standard written American English. The idiosyncratic collocations listed in this section have been judged to involve primarily grammatical irregularity; other contributing motivations will be discussed in chapter 5.

Grammatically-based idiosyncratic collocations can be divided into five types: prepositions, genitive structures, word order, word form, and mixed (which includes less common variations such as pronoun use and multiple variations). For example, a common idiosyncrasy is the overuse of the genitive structure to signal the modification of a noun. One student wrote **real consumers of women* instead of *real women consumers*. Another used **a characteristic of distinctive* in place of *a distinctive characteristic*. Word form irregularities (i.e., lacking a verb or having the wrong form of an adjective, noun, or verb) are also very common. For example, **worth to pay* appeared in one student's paper instead of the collocation *worth paying*, and **as can been seen* was used by several students in place of *as can be seen*. Another type of grammatical irregularity involved preposition use. Variants such as **to cater for popular tastes* and **do not mention about* constituted the 39 idiosyncratic collocations in this category. Each of the categories of grammatical irregularities in collocation is explained below, and each includes a listing of all idiosyncratic collocations of that variation type found in the student corpus.

Word form idiosyncrasy occurred 40 times in the corpus, making this the most productive category. All the learner collocations in this category involve the use of an

unusual adjective, noun, or verb form. Table 19 lists all collocations involving an idiosyncrasy in word form, along with the code for the student (ST) who wrote each one; the number of times a particular phrase occurred is indicated in parentheses. Irregular word forms are indicated in bold type.

TABLE 19

IDIOSYNCRATIC COLLOCATION INVOLVING WORD FORM

ST	Student Phrase	Target Collocation
WI (2)	As can been seen	As can be seen + has been seen
LU (3)	As can been seen	As can be seen + has been seen
PO (3)	As can been seen	As can be seen + has been seen
GO (3)	As could be seen	As can be seen
CH	jobbing retaining	job retention
CH	harmony working environment	harmonious working environment
PO (3)	price strategy	pricing strategy
DI	take acts	take action
LU	food, clothes , and shelters	food, clothing, and shelter
CH (3)	moral and freedom	morality and freedom
LY	consuming behavior of students	consumer behavior of students
LY	student's consuming type	student consumer type
LY	student's consuming behavior	student's consumer behavior
LY	student's consuming behaviors	student consumer behavior
LY	economical progress	economic progress
LY	economical prosperity	economic prosperity
LY	economical stability	economic stability
LY (2)	economical trend	economic trends
WI (3)	special consumer potentiality	special consumer potential
WI	figure out its potentiality	figure out its potential
DI (2)	the businessmen market	the business market
PO	memberships of the plans	members of the plans
PO	manager the property	manage the property
CH	assistant its growth	assist its growth
LY	to avoid loss money	to avoid losing money

The second most productive grammatical irregularity involves the use or lack of use of prepositions. English prepositions are notoriously difficult for the NNS to master, and at the same time they are key components of many collocations. Table 20 lists all preposition idiosyncrasies found in the corpus (a total of 39) along with the code for the student (ST)

who wrote each one; the number of times a particular phrase occurred is indicated in parentheses. The idiosyncratic phrases in table 20 are most often missing a preposition, though those placed at the bottom of table 20 contain the wrong preposition, indicated by bold text.

TABLE 20

IDIOSYNCRATIC COLLOCATION INVOLVING PREPOSITION

ST	Student Phrase	Target Collocation
PO	apply work	apply for work
GO (4)	care price	care about price
GO	care the price	care about price
CH	search jobs	search for jobs
CH	search the jobs	search for jobs
PO	the government appealed people	the government appealed to people
GO	To sum	To sum up
GO	adds its advertising budget	adds to its advertising budget
LU (2)	be close the nature	be close to nature
LU	to close the nature	be close to nature
CH	to close the nature	be close to nature
CH	to close nature	be close to nature
CH	to opt the concentrated marketing	to opt for concentrated marketing
LY	more dollars relative than	more dollars relative to
LU	complain their performances	complain about their performances
CH	feel tied about their finance	feel tied down by their finances
WI	points that	points out that
LY	to adapt the external climate	to adapt to the external climate
LY	to adapt the changing world	to adapt to the changing world
LY	To adapt the change	To adapt to the change
DI	to adapt these changes	to adapt to these changes
PO	to adapt the change	to adapt to the change
CH	to adapt the changing	to adapt to the changes
CH	to adapt the changing environment	to adapt to the changing environment
DI (2)	attention is paid on them	attention is paid to them
DI (2)	to go toward this direction	to go in this direction
DI	to cater for popular tastes	to cater to popular tastes
LY	key of achieving market success	key to achieving market success
LY	the key of the good strategies	the key to good strategies
LY	the key of success	the key to success
LY	pass through in front of us	pass by in front of us
LY	apply this theory into	apply this theory to
LY	target its new market into students	target students

The third most productive grammatical idiosyncrasy involves the use of a genitive structure instead of an adjectival modification. The idiosyncratic phrases in table 21 involve the use of a genitive structure where an adjectival construction is called for in most cases. Table 21 lists all genitive idiosyncrasies found in the corpus (a total of 32 collocations) along with the code for the student (ST) who wrote each one; the number of times a particular phrase occurred is indicated in parentheses. The problem structure is marked in bold type.

TABLE 21

IDIOSYNCRATIC COLLOCATION INVOLVING GENITIVE STRUCTURE

ST	Student Phrase	Target Collocation
CH	full of information's era	information-filled era
LU	lives' quality	quality of life
LU (3)	advertisements' budget	advertising budget
LU	analyze method	method of analysis
GO (2)	all styles o f works	all work styles
CH	because of the gap o f age	because of the age gap
LU (2)	budget o f advertising	advertising budget
WI (2)	real consumers o f women	real women consumers
WI	consumer o f women	women consumers
LU	war o f marketing	marketing war
CH	the change o f the demographic	changing demographics
CA (2)	the class o f business and working	business and working classes
PO (4)	the data o f consumer behavior	consumer behavior data
DI	the data o f statistic figures	statistical data + statistical figures
CA	the seriousness o f a jeopardy	serious jeopardy
CH (2)	conscious o f the moral	moral consciousness
CH	conscious o f moral	moral consciousness
CA	a characteristic o f distinctive	a distinctive characteristic
CA	a trend o f change	a changing trend
LU (3)	good performance o f sale	good sales performance

Word order idiosyncrasy occurred five times in the corpus, and involve incorrect word order in a modification structure. Table 22 lists all word order idiosyncrasy, along with the code for the student (ST) who wrote each one; the number of times a particular

phrase occurred is indicated in parentheses. Irregular word orders are indicated in bold type.

TABLE 22

IDIOSYNCRATIC COLLOCATION INVOLVING WORD ORDER

ST	Student Phrase	Target Collocation
CH	make them more worth than younger	make them worth more
CH (2)	older adults are less absent	older adults are absent less
CA (2)	securities backed on the mortgages	mortgage-backed securities

The last category of grammatical irregularities consists of 22 cases of either idiosyncrasy that could not be placed into one of the classes above, or contained multiple types of idiosyncrasy. Table 23 lists all these learner collocations, along with the code for the student (ST) who wrote each one; the number of times a particular phrase occurred is indicated in parentheses. Problem words are indicated in bold type.

TABLE 23

IDIOSYNCRATIC COLLOCATION INVOLVING MIXED VARIATION

ST	Student Phrase	Target Collocation
LU (4)	the degree of how	the degree to which
CH	the degree of how	the degree to which
LU	the degree to which how	the degree to which
LU (2)	the level of how	the level at which
PO (2)	the act in law	an act of law
GO	accomplish difficultly	accomplish with difficulty
CH	have very poor shapes	are in very poor shape
PO	worth to pay	worth paying
CH	worthy to do it	worth doing
CH	we do some survey to them	we survey them
GO	it is to noted that	it is to be noted that
LY	easy to aware	easily become aware of
CH	familiar the new skill	become familiar with the new skill
GO	under conditions	under these conditions
CA	Mostly important	Most importantly
CA	keep their eyes for	keep their eyes open for
GO	goods with low prices	low-priced goods

A summary of the number and type of grammatical idiosyncrasy for each student is shown in table 24.

TABLE 24

IDIOSYNCRATIC COLLOCATION INVOLVING PRIMARILY GRAMMAR

ST	Genitive	Order	Preposition	Form	Mixed	Total
CA	5	2	0	0	2	9
WI	3	0	1	6	0	10
LY	0	0	10	10	1	21
GO	2	0	7	3	4	16
LU	11	0	4	4	7	26
DI	1	0	6	3	0	10
PO	4	0	3	8	3	18
CH	6	3	8	6	5	28
Total	32	5	39	40	22	138

Table 25 contains a summary of the grammatical idiosyncrasy recorded in the corpus, according to the number of variation made by each student in each writing task, all of which are listed and explained in chapter 3. Each task is designated "in" for work written in class, or "out" for work written outside of class. Dashed lines in a cell represent papers that were not handed in. Rows in bold text represent an assignment which elicited ten or more idiosyncratic collocations involving grammatical irregularity. A total of 138 learner

collocations in which the most salient feature was grammatical irregularity were recorded over the entire student-produced corpus.

Following the learner collocation summary table is table 26, displaying the number of idiosyncrasies per one hundred words for each assignment. Cells with 0.5 or more learner collocations per 100 are indicated in bold.

TABLE 25
GRAMMAR COLLOCATION SUMMARY

Task	CA	WI	LY	GO	LU	DI	PO	CH	Total
1 in	0	--	--	--	1	0	0	0	1
2 in	0	--	--	--	0	--	0	0	0
3 in	0	0	0	1	--	0	0	0	1
4 in	0	0	1	0	0	0	1	1	3
5 out	0	2	0	2	4	0	3	3	14
6 out	0	0	1	0	1	1	0	0	3
7 out	2	2	1	4	9	3	3	4	28
8 out	0	0	0	--	1	1	0	0	2
9 out	1	0	1	--	0	1	1	1	5
10 out	--	0	2	0	0	0	1	1	4
11 out	0	0	0	0	0	0	0	0	0
12 out	0	0	0	0	0	0	0	1	1
13 out	0	0	0	--	0	--	--	0	0
14 out	0	1	1	0	0	--	0	2	4
15 out	2	4	5	3	8	2	3	2	29
16 out	3	0	5	1	0	1	3	5	18
17 out	0	1	2	0	1	0	2	1	7
18 in	0	0	0	2	1	0	0	0	3
19 out	1	0	2	2	0	1	1	6	13
Total	9	10	21	16	26	10	18	28	138

TABLE 26
GRAMMAR COLLOCATIONS PER 100 WORDS

Task	CA	WI	LY	GO	LU	DI	PO	CH	Ave
1 in	0	--	--	--	0.59	0	0	0	0.21
2 in	0	--	--	--	0	--	0	0	0
3 in	0	0	0	0.87	--	0	0	0	0.12
4 in	0	0	0.30	0	0	0	0.54	0.34	0.14
5 out	0	0.40	0	0.41	0.86	0	0.39	0.39	0.30
6 out	0	0	0.53	0	0.24	0.39	0	0	0.14
7 out	0.18	0.26	0.13	0.67	0.90	0.53	0.30	0.33	0.40
8 out	0	0	0	--	0.76	0.68	0	0	0.12
9 out	0.42	0	0.33	--	0	0.84	0.33	0.81	0.37
10 out	--	0	0.90	0	0	0	0.49	0.62	0.35
11 out	0	0	0	0	0	0	0	0	0
12 out	0	0	0	0	0	0	0	0.32	0.07
13 out	0	0	0	--	0	--	--	0	0
14 out	0	0.53	0.33	0	0	--	0	0.76	0.27
15 out	0.14	0.44	0.57	0.55	0.72	0.30	0.26	0.15	0.36
16 out	0.10	0	0.23	0.05	0	0.05	0.12	0.32	0.10
17 out	0	0.22	0.68	0	0.22	0	0.56	0.21	0.23
18 in	0	0	0	1.00	0.53	0	0	0	0.19
19 out	0.03	0	0.08	0.12	0	0.05	0.04	0.24	0.06
Total	0.07	0.09	0.21	0.21	0.28	0.14	0.17	0.28	0.17

4.4 Collocation idiosyncrasy involving semantic choice variation

A collocation is a domain-specific phrase in which one or more words express a very limited shade of meaning. An idiosyncratic collocation involving semantic variation entails the use of a near-synonym in place of the appropriate word for the given context; cases of variation involving obvious lexical transfer were not included in this category, as they are included in the lexical transfer category. For this corpus, idiosyncratic collocations involving semantic choice variation were more numerous than lexical transfer, phonological similarity, and grammatical irregularity combined.

Idiosyncratic collocations involving variation in semantic choice were categorized according to the Mel'čuk classification scheme outlined in chapter 3; additional lexical functions were taken from Frawley's (1988) adaptation of Mel'čuk where necessary. Each idiosyncratic word was placed in the category of the corresponding target collocation. For example, the phrase **forceful evidence* involves an incorrect choice to express a general idea of *very* in the context of the word *evidence*. It was therefore placed in the Magnification category, since that is where the corresponding target collocation *strong evidence* belongs.

Following is a brief explanation of each lexical function for which there is a corresponding semantic choice idiosyncrasy found in the corpus; all examples of each type found in the corpus are listed in a chart following each lexical function explanation. Each chart contains the student initials (ST), the student phrase (erroneous collocation) and the target collocation. Numbers in parentheses indicate the number of times a particular phrase was found in the corpus. Semantic choice variation are identified by bold type.

Oper_i (W₀) This function describes a semi-auxiliary verb, Oper, in which the role of W₀ is that of first complement in a sentence where the grammatical subject is the i-th

participant. In other words, this category contains a verb and its first following noun phrase. Mel'čuk and Zholkovsky (58) give the example, $\text{Oper}_1(\text{arrest}) = \text{make}$. From the corpus of student variation, an example is **make a meeting*, where $\text{Oper}_1(\text{meeting}) = \text{hold}$; thus, *hold a meeting*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
PO (2)	put an alternate for trustee	appoint an alternate for trustee
PO (2)	attending a legal service plan	joining a legal service plan
LY (2)	translating the dollars	converting dollars
GO (2)	contain the total number of hours	includes the total number of hours
GO	make a meeting	hold a meeting
CA	make a wrong definition	construct an incorrect definition
PO	makes no difference between	makes no distinction between
LY	replans its company goal	retargets its company goal
CH	to replace their jobs	to fill open positions
GO	assign time recording	enforce time-keeping
CA	carry lots of contributions	offer large contributions
PO	contact the topic of inheritance	encounter the issue of inheritance
PO	cultivating desalination plants	building desalination plants
PO	gather water	collect water
LU	receive new data	collect new data
CA	proceed a market poll	conduct a market poll
CH	change to diversified	become diversified
LU	to fit any kind of circumstances	to handle any kind of circumstances

Func_i (W₀) This function describes a second semi-auxiliary verb, Func, in which the role of W₀ is that of grammatical subject and the first surface complement is the i-th participant. Mel'čuk and Zholkovsky (58) give the example, $\text{Func}_1(\text{proposal}) = \text{be initiated}$ by someone. From the corpus of student phrases, an example is **the market can be diversified into several market segments*, where $\text{Func}_0(\text{market}) = \text{divide}$; thus, *the market can be divided into several market segments*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
WI	one command may be made	one command may be given
PO	the market can be diversified into	the market can be divided into
PO	the market was diversified in	the market can be divided into
PO (2)	cost which was saved	expense which was saved
LY	women occupy ... of consumers	women make up ... of consumers
LY	products are pointed	products are aimed at
CA	inclinations most students prefer	inclinations most students have
GO (2)	costs can contain	costs can include

Labor_{ij} (W₀) This function describes a third semi-auxiliary verb, Labor, in which the role of W₀ is that of the second surface complement. The grammatical subject is the i-th participant and the first surface complement is the j-th participant. Mel'čuk and Zholkovsky (59) give the example, Labor₁₂(interrogation) = subject, i.e., *X subjects someone to an interrogation*. From the corpus of student phrases, an example is **compile them into different categories*, where Labor₁₂(categories) = divide; thus, *divide them into different categories*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
LY	compile them into different categories	divide them into different categories

Real_i (W₀) This function describes Real, a verb analogous to the Oper function above, meaning 'to fulfill a demand contained in the meaning of W₀.' In this function, the role of W₀ is that of first complement in a sentence where the grammatical subject is the i-th participant. Mel'čuk and Zholkovsky (61) give the example, Real₂(hypothesis) = confirm [Facts confirm the hypothesis]. From the corpus of student phrases, an example is **make a meeting*, where Oper₁(meeting) = hold; thus, *hold a meeting*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
PO	takes it for example	provides an example of
LU	realize how the market is	discover the condition of the market
LU (2)	to prove my guess	to confirm my suspicions
CA	realize the reason	understand the reason
LY	rescue the food systems	conserve the food supply
LY	to accommodate the disaster	to prepare for the disaster

Liqu (W_0) This function, usually used in combination with verbal lexical functions (such as Func), describes a situation in which something is liquidated, or an action which results in a situation stopping or ceasing to occur. Here Liqu will be combined with Func to account for situations ceasing to occur. Mel'čuk and Zholkovsky (61) give the example, $LiquFunc_0$ (illiteracy) = wipe out [illiteracy]. From the corpus of student phrases, an example is **work out poverty*, where $LiquFunc_0$ (poverty) = stamp out; thus, *stamp out poverty*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
CA	work out poverty	stamp out poverty
GO	to clarify these doubts	to resolve these doubts
CH	to demise the conflicts	to end the conflicts
PO	extensively curtail ... control	rigidly curtail ... control
LU	largely decreased	largely eliminated

Note that the third phrase listed, **demise the conflicts*, could also be seen as a transitivity error.

Perm (W_0) This function, usually used in combination with verbal lexical functions, describes a situation which is allowed or permitted, or an action which results in a situation being allowed to occur. Mel'čuk and Zholkovsky (61) give the example, $PermOper_2$ (exam) = allow [someone to take an exam]. From the corpus of student phrases, an example is, **accept any interrupts*, where $PermOper_2$ (interruptions) = tolerate; thus, *tolerate interruptions*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
CH	accept any interrupts	tolerate interruptions
LY	developing space	room for development
GO	the chances of learning knowledge	opportunities to learn
PO	to survival the growing population	to save the growing population
PO	to survival their lives	to save their lives
CA	support this standpoint	support this viewpoint
DI	save the situation from worsening	keep the situation from worsening

Mult (W_0) This function describes an aggregate of a noun. Mel'čuk and Zholkovsky (56) give the example, Mult (student) = student body. From the corpus of student phrases, an example is **a number of evidence*, where Mult (evidence) = wealth; thus, *a wealth of evidence*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
CH	increasing number of older generation	increasing size of the older generation
LU	a number of evidence	a wealth of evidence
WI	a great number of money	a great deal of money
GO	a number of money	a lot of money
CH (3)	our middle life	our middle-age years
DI (2)	a large part of the people	a large number of people
LY	grand amount of currency exchange	large amount of currency exchange
LY	large amount of dollars	large amount of money
LY	spend so many time	spend so much time
CH	the rate of the consumers	the number of consumers
CH	they still own numerous knowledge	they still have a vast knowledge
PO (2)	keep the whole assets	keep all the assets
CH (3)	abundant working experience	a wealth of working experience
LU	in the rest year	in the other years
LU (2)	the portion of female user	the percentage of female users
LU	requirements including housing , clothing and eating	necessities including food, clothing and shelter

Magn (W_0) This function describes magnification, as in *very*. Mel'čuk and Zholkovsky (57) give the example, Magn (temperature) = high. From the corpus of student phrases, an example is **an intense impression*, where Magn (impression) = strong; thus, *a strong impression*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
WI	the more possible	the greater the possibility
GO (2)	to pay more salaries to employees	to pay higher salaries to employees
LU (2)	pay more salary	pay higher salaries
LU (3)	more market share	larger market share
DI	huge different action	largely different action
PO	huge abilities	great abilities
GO	so huge abilities	such great advantages
CA (2)	deeper status within each market	better/higher status
CH	their too cautious	their extreme caution
WI	in the unknown future	in the distant future
LY	long options for hedging	long-term options for hedging
DI	the aging generation	the older generation
CH	the expedition of the human life	increased life expectancy
CA	the higher desires	higher-level desires
DI	to lack well support	to lack good/strong support
DI	an intense impression	a strong impression
CA	forceful evidence	strong evidence
WI	extremely possible	extremely likely
WI (2)	greatly possible	extremely likely
GO (2)	fully important	extremely important
WI	enlarge the distance	increase the distance
WI	widen his knowledge	broaden his knowledge
WI (3)	accepted broadly	widely accepted

Incep (W₀) This function represents a verb expressing an initial stage of progression, and sometimes occurs together with verbal auxiliaries such as Oper. Mel'čuk and Zholkovsky (60) give the example, IncepOper₂ (rule) = fall [under the rule of]. From the corpus of student phrases, there is only one example: for the phrase, **a scientist raised a technique*, Incep (technique) = introduce, so that the native speaker collocation is *a scientist introduced a technique*. Following is the phrase which falls into this category:

ST	Student Phrase	Target Collocation
DI	a scientist raised a technique	a scientist introduced a technique

Cont (W₀) This function represents a verb expressing a continuing action. Mel'čuk and Zholkovsky (60) give the example, ContOper₂ (influence) = remain [under the influence of]. From the corpus of student phrases, for the phrase, **demands are still*

remained, Cont (same) = remain, so that the correct collocation is *demand remains the same*. This could also be seen as a passivization problem. Following are the two phrases which fall into this category:

ST	Student Phrase	Target Collocation
CH	catch up the step of the changing world	keep in step with
CH	catch up with the step of	keep in step with

Culm (W₀) This function represents the culmination or climax of a situation.

Mel'čuk and Zholkovsky (56) focus on the location aspect of this function, and they label it "Centr." They give the example, Centr (glory) = summit [of glory]. For the purposes of this project, this function is labeled Culm (W₀). From the corpus of student phrases, for the phrase, **attitudes must be met*, Culm (expectation) = met, so that the correct collocation is *expectations must be met*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
PO	fit the desires of them	meet their needs
PO (3)	fit the ... needs	meet the ... needs
LU	to fit the need	to meet the need
CH	satisfy their value	satisfy their desire
LU	satisfy what they need	meet their needs
LU	satisfy what the employers need	meet the employers' need
CA	content the basic needs	satisfy the basic needs
LU	crowded with the contradictions	filled with contradictions
CH (3)	gratify what they think	gratify their desires
DI (2)	matches consumers needs	meets consumers needs
LY	the rescue for the morality crisis	the solution for the morality crisis
DI (2)	win the success	achieve success
LU	completely realize	completely/fully understand
CA (2)	fully realize	completely/fully understand
WI	solved thoroughly	solved completely
CA	satisfaction of necessity	satisfaction of needs
CA (2)	full satisfaction	complete satisfaction
PO	attitudes must be met	expectations must be met

Excess (W_0) This function represents a situation in which something is functioning excessively. Frawley (365) gives the example, Excess (eyelid) = flutter. From the corpus of student phrases, for the phrase, **supply has been over demand*, Excess (supply) = exceed; thus, the native speaker collocation is *supply has exceeded demand*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
GO	work in hot time	work during the busiest time/the rush
GO	hot working time	busiest time (rush hour/the rush)
CH (2)	the supply have been over demand	supply exceeded demand
CH	supply has been over demand	supply has exceeded demand

Manif (W_0) This function represents a situation in which something becomes manifest or apparent. Mel'čuk and Zholkovsky (62) give the example, Manif (guilt) = become apparent. From the corpus of student phrases, for the phrase, **the role they act*, Manif (role) = play; thus, the native speaker collocation is *the role they play*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
CA	act a role like	play the role of + act like
CA	which roles do we want to act	which roles do we want to play
CA (2)	express specialty	express individuality
CA	express their distinctive flavors	express their distinctive tastes
DI (2)	the role they act	the role they play
LY	pose a significant role for	plays a significant role in

Minus (W_0) This function represents a lesser value or a situation in which something happens to a lesser degree. This is often used within a compound function. Frawley (367) gives the example, Minus (wind) = slacken. From the corpus of student phrases, for the phrase, **to save taxation*, Minus (taxation) = reduce; thus, the native

speaker collocation is *to reduce taxation*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
LY (2)	London's share is falling down	London's share is decreasing
LY	lower down the price	lower the price
LY	dollars keep falling down	the dollar keeps falling + going down
GO (2)	market share falls down	market share falls to + drops down
GO	market share has fallen down	market share has fallen
GO	market share has fallen down	market share has fallen
WI (2)	(sales volume) is still slipping down	(sales volume) is still slipping
PO (2)	the method of saving tax	the method of reducing/lowering tax
PO	saving taxation	reducing taxation
PO	to save taxation	to reduce taxation
PO (3)	save the income tax	reduce the income tax
DI	decrease the rate of crime	lower the rate of crime
WI	market segments decreased	market segments shrank
GO	losses of time cost	loss of time
GO	cost losses	financial losses
CH	declining the rate of sick	reducing the rate of illness
CH	retirees' health is accelerating failure	increasingly failing health
GO	stability is reducing	stability is declining
CA	sales are still monotonous	sales are still slow
WI	a little slipping down	slipping a bit
WI (4)	dramatically going down	dramatically decreasing /falling dramatically
CH	less responsibilities	fewer responsibilities

Plus (W_0) This function represents a greater value or a situation in which something happens to a greater degree. This is often used within a compound function. Frawley (367) gives the example, Plus (joy) = grow. From the corpus of student phrases, for the phrase, **heighten his quality of life*, Plus (quality of life) = improve; thus, the native speaker collocation is *improve his quality of life*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
CA (3)	raise the number of advertisements	increase the number of advertisements
CA	raising the number of advertisements	increasing the number ...
DI	Polio can raise its sales	Polio can increase its sales
DI	rise its sales	increase its sales
CA	attain many edges	gain an edge
LU	gain more competitive	become more competitive
GO	heighten his quality of life	improve his quality of life
CA	reinvigorate the market share	increase the market share
CA	retrieve old customers	regain former customers
GO (2)	suffered from ... improvement	benefited from ... improvement
CH	take a lot of advantages by	gain a lot of advantages by
CH	to grasp knowledge	to acquire knowledge
LY	build up their own features	build up their own reputations
LY	the backup of the greenback	the reemergence of the greenback
LU	to gain more competition	to gain a larger share of the market
LY	filled with explosive technology	filled with increasing technology
PO	by learning knowledge	by acquiring knowledge
WI	further realize	further understand
LY	push the US dollar's rising	push the US dollar upward

Obstr (W₀) A situation in which something is obstructed or functioning with difficulty. Frawley (367) gives the example, Obstr (justice) = obstruct. From the corpus of student phrases, for the phrase, **losing your friend's time*, Obstr (time) = waste; thus, the native speaker collocation is *wasting your friend's time*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
GO	out of the trend	out of step with current trends
CH	losing your friend's time	wasting your friend's time
GO	quit this prescription	refuse to take this prescription
WI	falses at the beginning	false starts
LY	On the opposite	On the contrary

Bon (W₀) A standard praise for the entry, W₀. Frawley (364) gives the example, Bon (advice) = sound. From the corpus of student phrases, for the phrase, **definite*

customers, Bon (customers) = loyal; thus, the native speaker collocation is *loyal customers*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
GO	ethic in value concepts	ethical values
GO	ethic value concepts	ethical values
CA (2)	value concepts	value judgments
WI (2)	value concepts	value judgments
DI (4)	value concepts	value judgments
WI	rooted concepts	deeply-rooted convictions
DI	definite customers	loyal customers
LY	faithful on the same products	loyal to the same products
CA (2)	original clients	loyal customers
LY (2)	high class cloth	expensive clothing
CA (2)	unique flavor	distinctive taste
CA	expensive objects	luxury items
CA	luxury objects	luxury items

Prox (W_0) This function represents proximity: something which is about to happen or which is on the verge of happening. Frawley (367) gives the example, Prox (disaster) = on the brink of. This category has been broadened to include all collocations having to do with periods of time, since they are not specifically covered in Mel'čuk's list of basic lexical relations. From the corpus of student phrases, for the phrase, **impending era*, Prox (years) = coming; thus, the native speaker collocation is *coming years*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
GO (5)	In past time	In the past
CA	impending era	coming years
PO	in a few years ago	a few years ago
LY	since six months ago	six months ago
PO	successful once in the past	successful in the past
DI (2)	nowadays large numbers	recent large numbers
CH	the left 35 years	the remaining 35 years
PO	original beginning of spenthrift trust	origin of spenthrift trust

$A_i (W_0)$ This function describes a typical adjective for the i -th participant in a given situation. This category holds only those adjective-noun phrases not described in the more specific categories above. Mel'čuk and Zholkovsky (56) give the example, A_1 (speed) = considerable [at a considerable speed]. From the corpus of student phrases, an example would be **in this rival society*, where A_1 (society) = competitive; thus, the native speaker collocation is *in this competitive society*. Following are the phrases which fall into this category:

ST	Student Phrase	Target Collocation
LU (2)	business portion	business segment
GO (3)	business class	business market/segment
GO (2)	business class to student class	business segment to student segment
CH (4)	market share rate	market share
LY	raise the crimes committing rate	raise the crime rate
GO (2)	working time	hours of operation
CH (2)	better health technique	better health habits
PO	complex works	complex tasks
LU	market characteristic	market trends
PO	the different education style	a different education system
DI (3)	the product figure	the product profile
LU (2)	pay practice	wage/pay scale
CH	a certain years	a certain age
DI (2)	purchasing abilities	purchasing power
CH	both physical and mind condition	both physical and mental condition
PO (3)	customer psychology	consumer psychology
LY	in this rival society	in this competitive society
LU	unbenefit behavior	dangerous behavior
GO (2)	attendance cards	time cards
CH	the average life of retirees	the average age of retirees
PO (2)	the routine law service	routine legal services
PO	the salary notice	the pay slip
CH	years of their rest life	years of their retired life
PO	to use their political rights	to use their political power
GO	egalitarian society to money society	egalitarian ... capitalistic society
PO	middle level persons	middle-class people
CH	the aging era	old age
CA (4)	initial customers	new customers
WI	consuming ability	consumption potential
LU	living space	living conditions
LY	portion for marketing	aspect of marketing

Adv_i (W₀) This function describes a typical qualification of an action for the i-th participant in a given situation, expressed by an adverbial. This category holds only those phrases not described in the more specific categories above. Mel'čuk and Zholkovsky (57) give the example, Adv₁ (speed) = at a speed of [tear along at a speed of 100 km an hour]. From the corpus of student phrases, the only example is **scan the two systems carefully*, where A₁ (examine) = carefully; thus, the native speaker collocation is *examine the two systems carefully*. Following is the phrase which falls into this category:

ST	Student Phrase	Target Collocation
GO	scan the two systems carefully	examine the two systems carefully

Table 27 contains a summary of the semantic choice learner collocations recorded in the corpus, according to the number of learner collocations produced by each student in each writing task. Rows in bold text represent an assignment which elicited ten or more idiosyncratic collocations involving semantic choice. A total of 292 learner collocations due to semantic choice were recorded over the entire student-produced corpus.

Following the learner collocation summary table is table 28, displaying the number of learner collocations per one hundred words for each assignment. Cells with 0.5 or more collocations per 100 words are indicated in bold.

TABLE 27
SEMANTIC CHOICE COLLOCATION SUMMARY

Task	CA	WI	LY	GO	LU	DI	PO	CH	Total
1 in	0	--	--	--	1	0	0	0	1
2 in	0	--	--	--	0	--	1	0	1
3 in	1	1	0	2	--	0	0	0	4
4 in	5	4	2	2	1	2	1	1	18
5 out	7	6	0	3	1	4	3	3	27
6 out	3	0	1	0	1	0	2	0	7
7 out	9	5	3	3	8	4	2	4	38
8 out	2	0	1	--	1	1	1	1	7
9 out	1	1	3	--	2	3	6	0	16
10 out	--	1	1	2	1	1	0	0	6
11 out	0	0	1	2	0	0	2	0	5
12 out	0	0	1	1	0	1	0	0	3
13 out	0	0	1	--	0	--	--	0	1
14 out	2	0	2	0	0	--	1	5	10
15 out	9	4	5	4	6	4	2	5	39
16 out	2	1	3	11	3	5	9	9	43
17 out	0	2	2	3	1	0	1	1	10
18 in	1	1	1	4	0	0	1	2	10
19 out	1	2	4	9	6	5	12	11	50
Total	43	28	31	46	32	30	44	42	292

TABLE 28
SEMANTIC CHOICE COLLOCATIONS PER 100 WORDS

Task	CA	WI	LY	GO	LU	DI	PO	CH	Ave
1 in	0	--	--	--	0.59	0	0	0	0.21
2 in	0	--	--	--	0	--	1.40	0	0.21
3 in	0.69	0.66	0	1.74	--	0	0	0	0.50
4 in	1.82	1.25	0.90	0.71	1.22	0.70	0.54	0.34	0.94
5 out	1.04	1.20	0	0.62	0.65	0.97	0.64	0.53	0.69
6 out	1.44	0.91	0.53	0	0.48	0	1.09	0	0.50
7 out	0.91	0.66	0.53	0.50	0.90	0.70	0.30	0.41	0.61
8 out	0.55	0	0.29	--	0.76	0	0	0.49	0.42
9 out	0.84	0.79	0.66	--	1.56	1.68	2.00	0	1.12
10 out	--	0.61	0.45	1.66	0.57	2.06	0	0	0.61
11 out	0	0	0.39	1.06	0	0	0.93	0	0.31
12 out	0	0	0.42	0.79	0	1.12	0	0	0.21
13 out	0	0	0.39	--	0	--	--	0.31	0.18
14 out	1.30	0	0.66	0	0	--	0.44	1.91	0.75
15 out	0.62	0.56	0.68	0.73	0.99	0.61	0.26	0.46	0.60
16 out	0.10	0.04	0.19	0.54	0.23	0.27	0.41	0.70	0.28
17 out	0	0.44	0.68	1.43	0.22	0	0.28	0.21	0.37
18 in	0.44	1.84	1.03	2.02	0	0	0.54	0.88	0.83
19 out	0.06	0.08	0.20	0.65	0.26	0.23	0.48	0.44	0.27
Total	0.38	0.30	0.37	0.69	0.47	0.41	0.48	0.46	0.43

Table 29 contains a summary of all learner collocations recorded in the corpus, according to the number of learner collocations produced by each student in each writing task, all of which are listed and explained in chapter 3. Each task is designated "in" for work written in class, or "out" for work written outside of class. Dashed lines in a cell represent papers that were not handed in. Rows in bold text represent an assignment which elicited ten or more idiosyncratic collocations involving semantic choice. A total of 571 learner collocations were recorded over the entire student-produced corpus.

Following the learner collocation summary table is table 30, displaying the total number of learner collocations per one hundred words for each assignment. Cells with 0.5 or more learner collocations per 100 words are indicated in bold. Note that overall, only one student, Willy, had a total learner collocation count of less than 0.5 learner collocations per 100 words.

TABLE 29
TOTAL COLLOCATION SUMMARY

Task	CA	WI	LY	GO	LU	DI	PO	CH	Total
1 in	0	--	--	--	2	0	0	0	2
2 in	0	--	--	--	0	--	1	0	1
3 in	2	2	1	3	--	0	0	0	8
4 in	6	4	5	3	2	2	2	3	27
5 out	9	9	5	8	9	5	12	11	68
6 out	3	2	2	0	4	1	3	0	15
7 out	18	10	9	8	21	8	8	10	92
8 out	2	0	1	--	2	2	1	1	9
9 out	3	2	4	--	2	4	9	3	27
10 out	--	1	4	2	1	2	1	3	14
11 out	0	0	1	3	0	0	2	0	6
12 out	2	0	1	1	0	1	0	2	7
13 out	0	0	1	--	0	--	--	3	4
14 out	3	0	6	0	0	--	1	8	18
15 out	16	9	19	7	22	7	8	10	98
16 out	14	3	16	14	6	9	15	23	100
17 out	0	2	4	5	2	0	3	2	18
18 in	1	3	3	6	3	0	2	2	20
19 out	8	5	10	15	8	10	15	22	93
Total	87	52	92	75	84	51	83	105	571

TABLE 30
TOTAL COLLOCATIONS PER 100 WORDS

Task	CA	WI	LY	GO	LU	DI	PO	CH	Ave
1 in	0	--	--	--	1.18	0	0	0	0.42
2 in	0	--	--	--	0	--	1.40	1.61	0.21
3 in	1.40	1.32	1.16	2.61	--	0	0	0	1.00
4 in	2.20	1.25	1.50	1.07	1.22	0.70	1.09	1.03	1.27
5 out	1.34	1.79	0.93	1.65	1.94	1.21	1.55	1.45	1.48
6 out	1.44	0.91	1.05	0	0.95	0.39	1.09	0	0.68
7 out	1.64	1.31	1.20	1.34	2.09	1.40	0.80	0.81	1.31
8 out	0.55	0	0.29	--	1.52	0.68	0	0.49	0.54
9 out	1.27	1.59	1.32	--	1.56	3.36	3.00	2.42	2.02
10 out	--	0.61	1.81	1.67	0.57	2.06	0.49	1.86	1.23
11 out	0	0	0.39	1.59	0	0	0.93	0	0.38
12 out	1.05	0	0.42	0.79	0	1.12	0	0.64	0.49
13 out	0	0	0.39	--	0	--	--	0.93	0.36
14 out	1.30	0	1.99	0	0	--	0.44	3.05	1.23
15 out	1.11	1.00	2.17	1.27	1.97	1.06	0.69	0.77	1.23
16 out	0.48	0.11	0.74	0.75	0.35	0.49	0.62	1.47	0.58
17 out	0	0.44	1.36	1.79	0.44	0	0.85	0.43	0.60
18 in	0.44	1.84	1.54	3.03	1.60	0	1.08	0.88	1.28
19 out	0.24	0.13	0.40	0.89	0.35	0.46	0.56	0.89	0.44
Total	0.69	0.46	0.94	1.07	0.92	0.70	0.79	1.00	0.80

4.5 Residue

The learner collocations in table 31 were labeled as residue, either because they may not actually be collocations, or they do not clearly fit into one of the categories described above.

TABLE 31
COLLOCATION RESIDUE

ST	Student Phrase	Target Collocation
CA	reinvigorate the withering marketshare	increase the shrinking marketshare
CA (2)	With expectation	as expected
CA	dead in water	dead in the water
LY	potential possibility of developing	potential/possibility of developing
GO (3)	record of attendance	time sheet
LU	to know how their potentials are	to know what their potential wants are
LU (2)	freedom as the first position	freedom first
DI	human beings have potential abilities	human beings have the potential
DI	it has not been clearly ensured	it has not been clearly determined
PO (2)	as a result of doing justice	as a result of a lawsuit
CA	paves a way	paves the way

4.6 Conclusions

The data displayed in this chapter reveal that collocation was a consistent problem for the students who participated in this study, across all four major types. In particular, the semantic category of collocational idiosyncrasy is most pervasive. Chapter 5 will explore the nature of the learner collocations found in the corpus.

CHAPTER 5

DISCUSSION OF THE FINDINGS

5.0 Introduction

This chapter will propose an analysis of the data presented in chapter 4, with the goal of characterizing the learner collocations examined in this study. Three main sections comprise chapter 5: an initial analysis of the data into two major types (section 5.1), and a deeper analysis of each of those two types (sections 5.2 and 5.3). The chapter concludes with a tentative characterization of the idiosyncratic collocations under scrutiny here, and an analysis of when to anticipate these idiosyncrasies.

The initial categorization of idiosyncratic collocations, as discussed in chapter 3, resulted in four categories of NNS collocations, based purely on surface features: those missing a word, those appearing to be a blend of two collocations, those involving word order or grammatical role problems, and those involving semantic misappropriation (see Lombard 1995). These categories, though mutually exclusive, failed to take into account two aspects of idiosyncratic collocations that upon further inspection are key factors in an overall characterization. Those two factors are native language lexical transfer and phonology.

Accordingly, chapter 4 presents the results of the revised analysis of the corpus for idiosyncratic collocation. The collocations were divided into four categories: those involving lexical transfer from the native language; those involving phonological irregularity; those involving grammatical irregularity; and those involving semantic

irregularity. The seventeen idiosyncratic collocations that did not fit into the above categories necessitated the creation of a fifth category: residue.

The numbers presented in chapter 4 reveal several facts about collocational variation in the writing corpus. First, the most numerous type of idiosyncratic collocations are those involving semantic variation. Second, each student produced at least one idiosyncratic collocation in each category. Third, the data show which student had trouble with which collocations and when. The relevance of these facts, however, need to be tested to determine whether they form a possible basis for a characterization of the NNS collocation. In this chapter, evidence supporting such a characterization is sought through further scrutiny of the idiosyncratic collocation categories themselves.

5.1. Initial analysis of observed idiosyncratic collocations

One possible method of examining the data collected in this study is to assume that the categories chosen in chapter 4 are independent of each other. Then, using a chi-square test, one can determine the probability that the four types of collocational variation presented above accurately represent a normally distributed population in which the four categories of idiosyncratic collocation vary independent of one another. The advantage to using this approach to the data is that either a positive or negative answer can be interpreted and give insight into the nature of the idiosyncratic collocations collected in this study. If these four categories do fit a model of independence, a different approach to the data is in order. If, however, there are indications that the four categories are not behaving independently, then a closer look at the categories may reveal something significant about the idiosyncratic collocations themselves.

As stated in chapter 4, though each of the idiosyncratic collocations was placed in only a single category, many could fit into two or more categories. A chi-square test

requires that all possible categorizations for each collocation be taken into account. Therefore, each of the four types of collocation have to be put in contingency tables matched against each of the other three types. In other words, the semantics category has to be put into three contingency tables opposite phonology, lexical transfer, and grammar; phonology has to be put into contingency tables opposite translation and grammar and so on, until all six possibilities are expressed in tables. See appendix D for a complete listing of the cross-classifications of collocations.

5.1.1 Test of tokens

The first chi-square test was run on all 571 recorded data points (tokens), regardless of how many times a particular idiosyncratic collocation occurred in the corpus. Since there are four categories of idiosyncratic collocations, a set of six contingency tables was constructed to test all possible combinations. The null hypothesis is that the four types of idiosyncratic collocations represent a model of independence. The degrees of freedom for a two-by-two contingency table is one. The critical value above which the null hypothesis has a 99 percent chance of being rejected is 6.64 (i.e., χ^2 is greater than 6.64). Tables 32 through 37 display the observed frequencies, expected frequencies, and deviance for each of the six contingencies.

TABLE 32
GRAMMAR VS. PHONOLOGY TOKENS CONTINGENCY

(a) Observed Frequencies			
	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	36	137	173
Grammar -- N	70	328	398
Column Total	106	465	571

Table 32. *Continued.*

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	32.16	140.88	173
Grammar -- N	73.88	324.12	398
Column Total	106.00	465.00	571

(c) Deviance: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.36	0.14	0.49
Grammar -- N	0.26	0.04	0.30
Column Total	0.62	0.17	0.79

All four of the actual frequencies for the grammar versus phonology contingency are very close to the expected values and the chi-square value is 0.79, which is below the criterion value of 6.64 indicating that the null hypothesis cannot be rejected. Thus there is most likely no dependency between these two types of collocational idiosyncrasy.

TABLE 33

SEMANTIC VS. LEXICAL TRANSFER TOKENS CONTINGENCY

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	44	350	394
Semantic -- N	30	147	177
Column Total	74	497	571

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	51.06	342.94	394
Semantic -- N	22.94	154.06	177
Column Total	74.00	497.00	571

Table 33. *Continued.*

(c) Deviance: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	1.12	0.12	1.24
Semantic -- N	1.88	0.37	2.25
Column Total	3.00	0.49	3.49

The chi-square value in this case rounds off to 3.49. For table 33, the chi value is not large enough for the null hypothesis to be rejected. In other words, collocations coded for lexical transfer and semantic idiosyncrasy operate independently.

TABLE 34

PHONOLOGY VS. LEXICAL TRANSFER TOKENS CONTINGENCY

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	6	100	106
Phonology -- N	68	397	465
Column Total	74	497	571

(b) Expected Frequencies: $(\text{row total}) \times (\text{column total}) \div (\text{grand total})$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	13.74	92.26	106
Phonology -- N	60.26	404.74	465
Column Total	74.00	497.00	571

(c) Deviance: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	4.94	0.57	5.51
Phonology -- N	0.87	0.16	1.03
Column Total	5.81	0.73	6.54

The chi square value for the phonology vs. lexical transfer pair is 6.54, which is not significant at the $p < 0.01$ level. A closer look at the deviance for table 34 reveals that one cell is contributing over three-quarters of the total chi-square value. This is because the

observed number of cases in which phonology and lexical transfer are both a factor in an idiosyncratic collocation is only half of what is expected according to the model (only 6 exist in the database). The trend in table 34 is for either lexical transfer or phonology to be a factor. So, while the total deviance value indicates that this pair does not follow a model of independence, it is largely because of a single cell, a single distribution of collocations involving both lexical transfer and phonology.

TABLE 35
PHONOLOGY VS. SEMANTIC TOKENS CONTINGENCY

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	54	52	106
Phonology -- N	340	125	465
Column Total	394	177	571

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	73.14	32.86	106
Phonology -- N	320.86	144.14	465
Column Total	394.00	177.00	571

(c) Deviance: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	5.28	10.57	15.85
Phonology -- N	1.08	2.68	3.76
Column Total	6.36	13.25	19.61

The chi square value for the phonology vs. semantics pair is 19.61. This is above the critical value (6.64), which means that the probability that the semantic and phonology categories are independent is very low. However, according to the deviance in table 35 this is due to large deviance in two cells, both involving cases where the phonology category is

present. In particular, the observed frequency for the +phonology +semantic cell (54) is almost one-third less often than could be expected (73.14), resulting in a deviance of 5.28; the observed frequency for the +phonology -semantic cell (52) is almost two-thirds more often than could be expected (32.86), resulting in a deviance of 10.57. In addition to the deviation data, there is another interesting pattern in table 35. The two cells indicating +phonology -semantics and +semantics -phonology both have observed frequencies greater than their expected frequencies. This may indicate that these two are in a trade-off distribution, where idiosyncratic collocations tend to be either phonologically or semantically related.

TABLE 36
GRAMMAR VS. LEXICAL TRANSFER TOKENS CONTINGENCY

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	5	168	173
Grammar -- N	69	329	398
Column Total	74	497	571

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	22.42	150.58	173
Grammar -- N	51.58	346.42	398
Column Total	74.00	497.00	571

(c) Deviance: $[(\text{observed} - \text{expected}) - 0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	14.32	1.90	16.22
Grammar -- N	5.55	0.93	6.48
Column Total	19.87	2.83	22.70

The chi-square value for the grammar vs. lexical transfer contingency table is also significant. The calculated value 22.70 ($p < 0.01$) means that the null hypothesis may be rejected for grammar versus lexical transfer. Again, however, the total value is due in large part to the contributions of two cells. The observed frequency of collocations coded for both lexical transfer and grammatical variation are less than one-fourth of the expected frequency (only 5 in the database); cases in which a collocation is coded for lexical transfer but not for grammar is one-third higher than expected. The high chi-square value for this pair is the result of the tendency of idiosyncratic collocations to be distributed such that those involving lexical transfer also preclude grammatical involvement.

TABLE 37
GRAMMAR VS. SEMANTIC TOKENS CONTINGENCY

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	41	132	173
Grammar -- N	353	45	398
Column Total	394	177	571

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	119.37	53.63	173
Grammar -- N	274.63	123.37	398
Column Total	394.00	177.00	571

(c) Deviance: $[(\text{observed} - \text{expected}) - 0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	52.11	113.08	165.19
Grammar -- N	22.08	50.43	72.51
Column Total	74.19	163.51	237.70

The chi square for the grammar vs. semantics contingency table is 237.70, which means that the probability that these two types are appearing in a random pattern is extremely low. Every cell in the grammar vs. semantic deviance segment of table 37 is by itself significant. In fact, in the cell in which both a semantic and grammar element are present, the observed value is one-third the expected value; the same is true of the cell in which both elements are absent. Movement is in the direction of the cells involving one or the other of the elements: almost a third more collocations involving semantic influence and not grammar influence are observed than expected, while more than two times the number of collocations involving grammar influence and not semantic influence are observed than expected. This fact indicates that these two types may be strongly opposed.

There are several trends evident in the chi-square data shown above. First, the null hypothesis that the four categories of idiosyncratic collocations are acting like independent variables cannot be rejected for the grammar versus phonology pair, and the semantic versus lexical transfer pair. The null hypothesis may be rejected for the phonology versus semantic pair, the grammar versus lexical transfer pair, and the grammar versus semantic pair. The lexical transfer versus phonology pair is borderline, and will need further examination.

These trends, though interesting, need to be checked against the individual data to determine whether the same trends hold. Accordingly, all of the contingency pairs were checked again for each student. This resulted in some slight problems, such as the appearance of expected values below five in some cells, which will need to be taken into account during the interpretation of the tables. Table 38 displays the chi square values for each of the contingency pairs for each individual student in the study. The actual tables for these values appear in appendix E.

TABLE 38
 TOKEN CHI-SQUARE VALUES FOR EACH STUDENT UNDER EACH
 CONTINGENCY

	Tr./Sem.	Ph./Tr.	Ph./Sem.	Gr./Tr.	Gr./Sem.	Gr./Ph.
CA	9.60	0.95	0.35	3.33	20.07	0.95
CH	0.51	3.92	1.74	2.38	51.91	0.33
DI	6.10	2.20	2.73	4.25	24.64	5.93
GO	0.49	2.40	2.02	4.24	39.59	2.03
LU	2.71	2.23	1.84	0.81	25.17	1.14
LY	0.08	2.67	6.31	7.64	24.85	0.09
PO	0.60	2.17	6.11	2.92	39.94	1.20
WI	2.29	0.39	8.22	3.12	22.81	2.24

The strongest trend in table 38 is visible under the grammar versus semantic contingency pair (in bold type). All these values are significant, which is in line with the complete data set analysis that the null hypothesis may not hold true for this contingency. For the individual data, contingency tables for two of the students (CA and DI) contain one expected frequency cell with a value of less than five. However, for the data set taken as a whole, all individuals contribute to at least three cells in each of the six contingency tables.

5.1.2 Test of types

In order to verify the trends revealed in the analysis of the token data, a second set of chi-square tests was run on the data set, this time including only the types. That is, a new database was created in which the duplication of all idiosyncratic collocations was

eliminated, regardless of the identity of the author. The number of collocations in the types database is 414. Once again, a set of six contingency tables was constructed to test all possible combinations. Tables 39 through 44 display the observed frequencies, expected frequencies, and deviance for each of the six contingency tables.

TABLE 39
GRAMMAR VS. PHONOLOGY TYPES CONTINGENCY

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	23	95	118
Grammar -- N	54	242	296
Column Total	77	337	414

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	21.95	96.05	118
Grammar -- N	55.05	240.95	296
Column Total	77.00	337.00	414

(c) Deviance: $[(\text{observed} - \text{expected}) - 0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.01	0.03	0.04
Grammar -- N	0.04	0.00	0.04
Column Total	0.05	0.03	0.08

The chi-square value for grammar versus phonology types is 0.08, even lower than it was for the tokens data. All four cells are well below any significance value, which indicates again that the probability that these two are operating in a dependent relationship is less than 25 percent.

TABLE 40

SEMANTIC VS. LEXICAL TRANSFER TYPES CONTINGENCY

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	35	255	290
Semantic -- N	22	102	124
Column Total	57	357	414

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	39.93	250.07	290
Semantic -- N	17.07	106.93	124
Column Total	57.00	357.00	414

(c) Deviance: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.74	0.08	0.82
Semantic -- N	1.15	0.27	1.42
Column Total	1.89	0.35	2.24

For table 40, just as in the contingency table for tokens, the chi value is not large enough for the null hypothesis to be rejected. In other words, there is a 99 percent chance that for the given data set, collocations coded for lexical transfer and semantic idiosyncrasy operate independently.

TABLE 41

PHONOLOGY VS. LEXICAL TRANSFER TYPES CONTINGENCY

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	5	72	77
Phonology -- N	52	285	337
Column Total	57	357	414

Table 41. *Continued.*

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	10.60	66.40	77
Phonology -- N	46.40	290.60	337
Column Total	57.00	357.00	414

(c) Deviance: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	3.51	0.39	3.90
Phonology -- N	0.56	0.13	0.69
Column Total	4.07	0.52	4.59

The chi square value for the phonology vs. lexical transfer pair is 4.59, which is not significant at the $p < 0.01$ level. This follows the same trend as the token data for this contingency. The deviance in table 41 again shows that one cell is contributing over three-quarters of the chi-square total value. This is because the observed number of cases in which phonology and lexical transfer are both a factor in an idiosyncratic collocation is only half of what is expected according to the model (only 5 exist in the types database). The trend in table 41 is for those idiosyncratic collocations involving lexical transfer to not also involve phonology. So while the total deviance value indicates that this pair as a whole does not follow a model of independence, it is largely because of a single cell. The combination of lexical transfer and phonology as features in an idiosyncratic collocation is dis-favored.

TABLE 42

PHONOLOGY VS. SEMANTIC TYPES CONTINGENCY

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	42	35	77
Phonology -- N	248	89	337
Column Total	290	124	414

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	53.94	23.06	77
Phonology -- N	236.06	100.94	337
Column Total	290.00	124.00	414

(c) Deviance: $[(\text{observed} - \text{expected}) - 0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	2.87	5.67	8.54
Phonology -- N	0.55	1.54	2.09
Column Total	3.42	7.21	10.63

The chi square value for the phonology vs. semantics pair is 10.63, which, as in the contingency for tokens, is significant at the $p < 0.01$ level. Overall, the probability that the semantic and phonology categories are independent is very low. However, according to the deviance in table 42 this is due to a single large deviance in the cell where the phonology category is present and the semantic category is absent: there are fifty percent more in the observed cell than the expected cell. The reason for this is a movement from the cell in which both semantic and phonology involvement are present. The trade-off distribution postulated for the token data is supported by table 42.

TABLE 43

GRAMMAR VS. LEXICAL TRANSFER TYPES CONTINGENCY

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	3	115	118
Grammar -- N	54	242	296
Column Total	57	357	414

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	16.25	101.75	118
Grammar -- N	40.75	255.25	296
Column Total	57.00	357.00	414

(c) Deviance: $[(\text{observed} - \text{expected}) - 0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	11.63	1.60	13.23
Grammar -- N	3.99	0.73	4.72
Column Total	15.62	2.33	17.95

The calculated value for the grammar vs. lexical transfer types contingency is 17.95, indicating a high likelihood of a dependent relationship. Again, however, the total value is due in large part to the contributions of two cells. The observed frequency of collocations coded for both lexical transfer and grammatical idiosyncrasy are less than one-fourth of the expected frequency, just as was the case in the tokens contingency for this pair. Cases in which a collocation is coded for lexical transfer but not for grammar also parallels findings above: the observed frequency is one-third higher than expected. Overall, this contingency pair for types follows the same trends as the one performed for tokens: grammar and lexical transfer strongly tend not to co-occur as possible causes in idiosyncratic collocations.

TABLE 44

-- GRAMMAR VS. SEMANTICS TYPES CONTINGENCY

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	28	90	118
Grammar -- N	262	34	296
Column Total	290	124	414

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	82.66	35.34	118
Grammar -- N	207.34	88.66	296
Column Total	290.00	124.00	414

(c) Deviance: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	36.81	82.99	119.80
Grammar -- N	14.14	34.31	48.45
Column Total	50.95	117.30	168.25

The chi square for the grammar vs. semantics contingency for types is very similar to the one for tokens: the overall chi-square value is 168.25, and each single cell exceeds the criterion value. Movement again is in the direction of the cells involving one or the other of the elements: out of 414 total collocations in table 44, only 62 of them entail both or neither grammar and semantic involvement. This evidence supports the idea that these two types may be in a strongly oppositional relationship.

The trends evident in the chi-square tokens data are all supported in the types data. In particular, there is little reason to believe that there is anything but an independent relationship between the collocations in the grammar versus phonology pair and the semantic versus lexical transfer pair; there are definite trends indicating dependency between the phonology versus semantic pair, the grammar versus lexical transfer pair, and

the grammar versus semantic pair; the lexical transfer versus phonology pair is still questionable, but the types data indicate that it will be difficult to reject the null hypothesis without further interpretation or evidence.

While it is true that in this study the results for the analysis of type and token data were similar, this is not necessarily generalizable to other studies. In other words, it may be that one particular error is repeated by several students; using type data alone thus may obscure the results for certain categories of collocation, or cause unusual affects that would carry over from one individual's data into the entire data set. For this reason, before conclusions can be drawn concerning the trends discussed above, it is necessary to examine individual types data. After the database was revised to include all types for each individual (which raised the total number of types to 420), all of the contingency pairs were checked again for each student. Even more cells in this analysis had expected values below five due to the fact that all students had produced errors that occurred more than once in the corpus. Table 45 displays the chi square values for each of the contingency types pairs for each individual student in the study. The actual tables for these values appear in appendix F.

TABLE 45
TYPE CHI-SQUARE VALUES FOR EACH STUDENT UNDER EACH
CONTINGENCY

	Tr./Sem.	Ph./Tr.	Ph./Sem.	Gr./Tr.	Gr./Sem.	Gr./Ph.
CA	5.85	1.32	0.23	2.77	10.18	1.32
CH	0.56	3.19	0.58	2.31	39.46	0.11
DI	1.57	2.04	1.46	2.91	19.26	2.46
GO	0.32	2.31	2.03	3.33	24.91	2.04
LU	1.01	2.03	0.83	1.75	21.33	0.74
LY	0.24	2.22	5.94	7.05	25.05	0.07
PO	0.41	2.04	1.74	2.19	21.99	1.07
WI	2.08	0.44	2.68	2.38	16.75	0.77

The chi-square values for the individual types data is very similar to the individual tokens data. The strongest trend again appears under the grammar versus semantic contingency pair (in bold type). All these values are significant, which is in line with the complete data set analysis for the types data that the null hypothesis may not hold true for this contingency. Though not all expected frequency cells in the analysis of individual data contain values of five or more, all individuals contribute to at least three cells in each of the six contingency tables.

5.1.3 Conclusions concerning the Chi-square test on the four categories

5.1.3.1 Introduction

Section 5.1.3.2 will review the findings in 5.1.1 and 5.1.2. Section 5.1.3.3 will discuss the implications of those findings.

5.1.3.2 Review of the Chi-square findings

Two of the contingencies named above clearly fall within the limits of the null hypothesis (i.e., it is very unlikely that they do not fit a model of independence). Those two contingencies are the grammar versus phonology contingency and the lexical transfer versus semantics contingency. A third contingency, phonology versus semantic, failed to produce convincing evidence for the rejection of the null hypothesis; the probability of dependence was only 95 percent. Though these contingencies alone are not "significant" in statistical terms, they do contribute to the story told by the remaining contingency tables.

Table 46 summarizes the probabilities of dependence for each of the contingencies, based on the token data.

TABLE 46

P-VALUE SUMMARY FOR TOKEN CONTINGENCIES¹

	Ph.	Tr.	Sem.	Gr.
Ph.	---	0.01	> 0.01	0.37
Tr.	---	---	0.06	> 0.01
Sem.	---	---	---	> 0.01
Gr.	---	---	---	---

¹ Actual values: Phonology vs. Semantic: 9.5 E-06; Grammar vs. Transfer: 1.9 E-06; Grammar vs. Semantic: 1.3 E-53.

The most salient probability datum for the six contingencies tested is the grammar versus semantics contingency, which provides the strongest evidence for rejecting the null hypothesis. The reason for the consistently high chi-square values in this contingency is a clear distribution of the two categories away from one another, a resistance to overlap in contributing to idiosyncratic collocations. Evidence for this resistance can be found throughout the data. For example, **quit this prescription* is a perfectly grammatical phrase for standard American written English; the problem is that the semantic sense of *quit* doesn't fit the given context. The same can be said of the student phrases **a scientist raised a technique*, and **completely realize*. In the case of learner collocations involving grammatical idiosyncrasies, students generally managed to produce at least the root content words appropriate to the context, as in **to cater for popular tastes*, **budget of advertising*, and **food, clothes, and shelters*.

Other contingencies display this resistance to overlap as well: the grammar versus lexical transfer and the semantic versus phonology contingencies exhibit high chi-square values. In the case of the grammar versus lexical transfer contingency, the significant skewing is from the cell involving both categories to the cell involving only lexical transfer (i.e., the collocations involving lexical transfer tend not to also involve grammar). In the case of the semantic versus phonology contingency, movement is within the phonology category from the cell involving both phonological and semantic influence to the cell involving only phonological influence.

5.1.3.3 Implications of the findings

A high probability value under the conditions outlined above indicates that the category pairs are not independent but pattern significantly with one another. If so, the two categories patterning in this way are likely to be more simply expressed by a single category. The semantic and grammar pair have a consistently high chi-square value, and are

thus the best candidates for reduction into a single category. Since the semantic category is the largest in the data set (involving 69 percent of the data set), it is assumed to be an important factor in the production of idiosyncratic collocations. Therefore, the distribution of collocations with respect to semantics will be used to describe a model of overall distribution of collocations. Each of the other categories must add to the model, or they are simply empty categories. If the grammar category is a mirror image of the semantic category in terms of always involving only a structural idiosyncrasy while the semantic category always involves only a meaning idiosyncrasy, including grammar will add nothing to the model. A close look at the grammar versus semantic contingency table for actual frequency reveals that the cells involving the presence of a grammatical element express a reversal of the expected semantic frequencies. Instead of the expected 69 percent "yes" and 31 percent "no," table 37 shows a 26 percent "yes" and a 74 percent "no." This means that the grammar dimension does add something to the model and therefore cannot be reduced into the semantic category.

While the lexical transfer and semantic categories and the grammar and phonology categories do not express high chi-square values, there is other evidence that these pairs may be reducible. The lexical transfer and semantic categories are alike in that they often co-occur and they share a resistance to overlap with the grammar category. The grammar and phonology categories are alike in that they often co-occur and they share a resistance to overlap with the semantic category. These two pairs of categories, then, are each more simply expressed by one. In other words, the phonology and grammar categories (those concerned with form) can be reduced to one category; likewise, the semantic and lexical transfer categories (those concerned with meaning) can be reduced to one. This would mean that the idiosyncratic collocations found in this data tend to be either form or meaning related.

This analysis and the resulting characteristic of idiosyncratic collocations to be irregular due to either form or meaning but seldom both constraints leads to more questions: Why does this opposition exist in this data? What does it mean? How can the collocation data be characterized along the deviance discussed above? Answers to these questions lie in a further analysis of the two opposed categories. This further analysis will rely on inspection of only data tokens (i.e., without a reduction to types) for two reasons. First, the previous analysis of the four data categories turned up quite similar results for type and token tests. Second, the narrowing of the focus to two categories decreases the overall number of data points; further compression of the data may result in a loss of important insight into the character of individual collocations. Thus, sections 5.2 and 5.3 pursue a further analysis of the form-versus-meaning opposition present in the data by examining the two most significant categories: learner collocations involving grammar irregularity (section 5.2), and learner collocations involving semantic irregularity (section 5.3).

5.2 Examination of idiosyncrasies involving grammar

In chapter 4, idiosyncratic collocations involving grammar were divided into five types based upon the structural elements present in those collocations: prepositions, genitive structures, word order, word form, and mixed (those not fitting into the other four classes). Of these five types, four involve word-external structural variation; the other (word form) involves word-internal structural variation.

The first step in further analyzing these types is to test a generalization about them in terms of a model. In producing a collocation with idiosyncratic structural form, a student is either relying on a native-language structure or is relying on some other information to make a decision about structure. An interesting generalization to test is whether the

idiosyncrasies in form are more likely to arise from a reliance on the native language or a reliance on some other factor(s).

In this study, students were native speakers of Mandarin, which, according to Li and Thompson (1981), should be classified as an isolating language. Each word consists of one morpheme, and there are very few inflectional morphemes in the language. The result is that Mandarin and English have some striking differences on the morpheme level. For example, in Mandarin the plural is only marked in special cases, such as on pronouns; aspect and tense of verbs are often marked at the end of the sentence, if at all; nouns, verbs, adjectives, and adverbs do not change form to "agree" with other members of a sentence. When native speakers of Mandarin produce collocations containing idiosyncratic word forms (e.g., *jobbing retaining), it is very unlikely that they could be relying on a native-language structure in order to produce them. In deciding on the form of an English word, then, native speakers of Mandarin have no choice but to rely on something other than their knowledge of their native language.

However, in the case of the four types of idiosyncratic collocations involving word-external structural variation (i.e., prepositions genitive structure, word order, and mixed), is there some pattern or prediction that can be made about which of the two strategies students are more likely to rely on? This is the question that will be addressed in this section.

A chi-square test was again used to measure the goodness of fit of the generalization explained above (i.e., that irregularity in form is due to the learner's reliance on either native language or reliance on something else) to a model of independence, across the five variation types. Each of the idiosyncratic collocations in the grammar category was marked as due to native language structure (Native Structure) or some other strategy (Unknown), based on the researcher's knowledge of the structure of Chinese, with the

assistance of a native speaker. The null hypothesis is that these two causes will vary independent of each other.

The word order and mixed categories were combined into an "other" category, since neither alone contained enough data points to ensure an expected value of at least five in each cell. Nevertheless, the expected frequency of Native Structure tokens in the "other" category is less than five. According to Woods, Fletcher, and Hughes (1986: 145), small expected frequencies is a potential problem for a chi-square test. However, they add that if cells with small expected frequencies contribute only slightly to the total deviance, it is unlikely that the value has been distorted. If this is the case, the results of the test can be accepted.

Table 47 displays the data for the grammar analysis chi-square test.

TABLE 47

NATIVE STRUCTURE/UNKNOWN INFLUENCE VS. GRAMMAR CATEGORIES

(a) Observed Frequencies

Categories	Prep	Gen	Word Form	Other	Totals
Native Structure	13	8	0	1	22
Unknown	26	24	40	26	116
Totals	39	32	40	27	138

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

Categories	Prep	Gen	Word Form	Other	Totals
Native Structure	6.22	5.10	6.38	4.30	22
Unknown	32.78	26.90	33.62	22.70	116
Totals	39.00	32.00	40.00	27.00	138

(c) Deviance: (observed-expected)² ÷ expected

Categories	Prep	Gen	Word Form	Other	Totals
Native Structure	7.40	1.65	6.38	2.54	17.96
Unknown	1.40	0.31	1.21	0.48	3.41
Totals	8.80	1.96	7.59	3.02	21.37

The chi-square test for the grammar analysis has three degrees of freedom: the chi-square value is significant for $\chi^2 > 7.80$ at the $p < 0.05$ level, and for $\chi^2 > 11.31$ at the $p < 0.01$ level. Though the chi-square value above of 21.37 appears to be significant (and thus appears to support the reasoning that the null hypothesis cannot not be rejected), only an examination of individual cells can explain why this value is high. The deviance values for Preposition and Word Form for reliance on native language structure together comprise almost two thirds of the total chi-square value; the Word Form category for Native structure, suspect because of its low expected value, contributes less than one-seventh of the total deviance.

Overall, the frequency of native language structure is much less than that of the structures from an unknown source. For the Preposition category, there are almost twice as many observed occurrences of collocations in which students used a native language structure as could have been expected. For the Word Form category, there are zero observed collocations in which students relied on native language structure but 6.38 are expected. Given this strong tendency for collocations to involve the unknown, the fact that word form collocations cannot involve native language structure, and the tendency of prepositions to involve native language structures, the Other category might be expected to have a lower incidence of collocations involving native language structures.

In terms of word form, the high deviance value results from the fact that Chinese students have no Chinese structure to rely on in constructing English words: as explained above, the two languages are quite different in that respect. In terms of prepositions, the Chinese language uses fewer of them than does English. Of the 13 observed collocations in the native language structure preposition cell, eight were placed there by virtue of a missing preposition; five had a preposition which did not fit the context. Overall, these results indicate that when native speakers of Chinese have difficulty with the structure of a collocation in terms of preposition or word form, the causes are not distributed randomly.

In other words, an idiosyncrasy in collocational form due to preposition is more likely to be the result of a reliance on Chinese structure than a random distribution would predict. Likewise, an idiosyncrasy in collocational form due to word form is more likely to be the result of a reliance on some factor besides the native language than a random distribution would predict. This other factor could be the overgeneralization of an English word formation rule (i.e., adding -ing to a word to make a noun, as in **student's consuming type*; adding -al to a word to make it an adjective, as in **economical progress*), the underutilization of such rules (e.g., **price strategy*), or possibly the confusion of British and American English forms.

5.3 Examination of idiosyncrasies involving semantics

Native language collocations are typically broken down in one of two ways: either the members of the collocation are deemed equal because they contain only content words (often referred to as "lexical collocations"); or they are divided into a dominant word and a grammatical structure (often referred to as "grammatical collocations"). In this study, the idiosyncratic collocations have already been categorized so as to separate those in which the idiosyncrasy involved structure. Therefore, the collocations under scrutiny in this section might best be referred to as lexical collocations. One method of analyzing the internal features of lexical collocations is to break them down into two components: a headword (or entry word, in the case of Mel'čuk's dictionary) and a collocate (Mel'čuk's functor). The remainder of section 5.3 describes the process by which the learner collocations involving semantic idiosyncrasy were broken down and examined.

5.3.1 Organization of the data

Accordingly, all idiosyncratic collocations originally recorded as involving a semantic element were divided into headword and functor pieces (these are the collocations for which the researcher deemed the most salient idiosyncrasy was semantic). Of the 293 idiosyncratic collocations coded as involving a semantic idiosyncrasy, 266 of the target versions were judged to have a noun as the headword (e.g., *vast knowledge*, *middle-age years*, *tolerate interruptions*), 20 of the target versions were judged to have a verb as the headword (e.g., *widely accepted*, *largely eliminated*, *rigidly curtail*), and seven target versions were judged to have an adjective as the headword (e.g., *extremely likely*).

The collocations themselves were then coded for the presence of an idiosyncrasy in either the headword, the functor, or both. For example, in the idiosyncratic collocation, **cost losses*, the target collocation is *financial losses*. The headword in this case is *losses*; the idiosyncrasy lies in the word *cost*. However, in **distinctive flavors*, the idiosyncrasy lies in the headword, *flavors*, which replaces the target, *tastes*. An example of both headword and functor as idiosyncratic members of a collocation is the production of **greatly possible*, in place of the target *extremely likely*.

Among the entire set of tokens, a total of 99 idiosyncratic headwords and 227 idiosyncratic functors were identified among the 293 collocations involving semantic idiosyncrasy; 30 of the collocations contain both types of idiosyncrasy. Each student produced more functor idiosyncrasies than headword idiosyncrasies, and all but two students produced at least twice as many functor as headword idiosyncrasies. Table 48 shows how many tokens each student contributed to each category.

TABLE 48
 PRESENCE OF HEADWORD AND FUNCTOR IDIOSYNCRASY

Student	Functor	Headword	Totals
CA	33	16	49
CH	33	15	48
DI	20	9	29
GO	37	13	50
LU	21	14	35
LY	26	6	32
PO	39	12	51
WI	17	13	30
Totals	227	99	326

The goal of this section is to examine the relationship between idiosyncrasies in headwords and functors of the semantically-related collocations and Mel'čuk's 20 categories under which they are arranged. In order to test the goodness of fit of the data to a model of independence, it was necessary to make two adjustments to the raw data. First, since the idiosyncrasies of headwords and functors overlap, it was necessary to divide them into two separate tests counting the presence and absence of each (i.e., one test in which the presence and absence of idiosyncratic headwords are counted for each category, and a second test in which the presence and absence of idiosyncratic functors are counted for each category). Second, it was necessary to reduce the number of Mel'čuk categories so that all cells would have an expected frequency of five or greater.

Mel'čuk's categories were reduced from 20 to seven in the following manner. The first four categories, Oper, Func, Labor, and Real were placed together (labeled O/R in the

tables) because they are grouped together in Mel'čuk's categorization scheme as verbal auxiliaries; Magn, Mult, and Plus were placed together (labeled M/M/P) because they all involve the description of semantically greater increments; Culm and Excess were placed together (labeled C/Ex) because they are semantically linked in describing a maximum; Liqui, Obstr, and Minus were placed together (labeled L/O/M) since they all involve the description of a semantically reduced or mitigated situation. The remaining six small categories (Adv, Incep, Cont, Perm, Manif, and Prox) were grouped together, since none of these could stand alone and have an expected frequency of at least five for the headword chi-square test. This category was labeled Other. Two of the original categories, Bon and Adj, were not combined with any other categories.

5.3.2 Headword and functor token Chi-square tests

Once the categories were combined, chi-square values were calculated to examine whether the spread of headword and functor idiosyncrasies across Mel'čuk's categories was skewed in any way. In other words, the null hypothesis is that individual Mel'čuk categories have nothing to do with the presence or absence of headword or functor idiosyncrasies, that they are independent variables. The results of the chi-square test for the presence and absence of headword idiosyncrasies in the semantically-related collocations are shown in table 49. This test involved all tokens in the set.

TABLE 49

HEADWORD IDIOSYNCRASY VS. MEL'ČUK REDUCED CATEGORIES

(a) Observed Frequencies

	Bon	Adj	M/M/P	C/Ex	L/O/M	O/R	Other	Row Total
Hw -- Y	19	30	14	11	8	9	8	99
Hw -- N	2	21	63	21	32	31	24	194
Column Tot	21	51	77	32	40	40	32	293

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Bon	Adj	M/M/P	C/Ex	L/O/M	O/R	Other	Row Total
Hw -- Y	7.10	17.23	26.02	10.81	13.52	13.52	10.81	99
Hw -- N	13.90	33.77	50.98	21.19	26.48	26.48	21.19	194
Column Tot	21.00	51.00	77.00	32.00	40.00	40.00	32.00	293

(c) Deviance: (observed-expected)² ÷ expected

	Bon	Adj	M/M/P	C/Ex	L/O/M	O/R	Other	Row Total
Hw -- Y	19.97	9.46	5.55	0.003	2.25	1.51	0.73	39.48
Hw -- N	10.19	4.83	2.83	0.002	1.15	0.77	0.37	20.15
Column Tot	30.16	14.29	8.38	0.005	3.40	2.28	1.10	59.62

Degrees of Freedom = 6

$\chi^2 > 12.59$ $p = 0.05$

$\chi^2 > 16.81$ $p = 0.01$

In the chi-square test for the headwords, each of the expected frequencies is greater than five. The total deviance is 59.62, which is significant at the $p = 0.01$ level. However, more than half of that value is contributed by the column for Bon (30.16)—Mel'čuk's lexical function for "standard praise" for an entry. In fact, the observed and expected values for Bon are quite different: nineteen headword idiosyncrasies were observed, and only slightly more than seven were expected; only two collocations lacked idiosyncrasy in the headword, but almost fourteen of these were expected. The Bon category demonstrates an almost complete reversal of what is expected. There is a very strong tendency of collocations in this category to involve a headword idiosyncrasy. The Adj category also

contributes a large proportion of the chi-square value (14.29). The total deviance value for the Adj column alone is greater than the criterion value of 12.59 for all of table 49. As with the Bon category, there were more idiosyncratic headwords than could be expected in a random distribution.

Of the other five categories, only one (M/M/P) contributes measurably to the chi-square value (8.38). In this case, however, the high value is due to lower observed frequencies than expected for idiosyncratic headwords. In this case, the tendency was for the idiosyncrasy to not involve the headword.

Before drawing any definitive conclusions concerning headwords and the Mel'čuk categories, it is necessary to examine the results of the chi-square test for the functors. The results of the chi-square test for the presence and absence of functor idiosyncrasies in the semantically-related collocations are shown in table 50.

TABLE 50
FUNCTOR IDIOSYNCRASY VS. MEL'ČUK REDUCED CATEGORIES

(a) Observed Frequencies

	Bon	Adj	M/M/P	C/Ex	L/O/M	O/R	Other	Row Total
Func -- Y	9	27	72	22	33	37	27	227
Func -- N	12	24	5	10	7	3	5	66
Column Tot	21	51	77	32	40	40	32	293

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Bon	Adj	M/M/P	C/Ex	L/O/M	O/R	Other	Row Total
Func -- Y	16.27	39.51	59.66	24.80	30.99	30.99	24.79	227
Func -- N	4.73	11.49	17.34	7.21	9.01	9.01	7.21	66
Column Tot	21.00	51.00	77.00	32.00	40.00	40.00	32.00	293

Table 50. *Continued.*

(c) Deviance: $(\text{observed} - \text{expected})^2 \div \text{expected}$

	Bon	Adj	M/M/P	C/Ex	L/O/M	O/R	Other	Row Total
Func -- Y	3.25	3.96	2.55	0.31	0.13	1.17	0.20	11.57
Func -- N	11.17	13.63	8.79	1.08	0.45	4.01	0.68	39.80
Column Tot	14.42	17.59	11.34	1.39	0.58	5.18	0.88	51.37

Degrees of Freedom = 6

$\chi^2 > 12.59$ $p = 0.05$

$\chi^2 > 16.81$ $p = 0.01$

One of the expected frequencies in table 50 is less than five, Bon, which has an expected value of 4.73. According to Woods, Fletcher, and Hughes (1986: 145), small expected frequencies may cause the corresponding deviance cells to contain a value higher than they ought to. Therefore, the interpretation of the corresponding Bon deviance cell will have to be conservative.

The total deviance is 51.37, which is significant at the $p = 0.01$ level. However, almost 85% of that value is contributed by three of the seven columns: Bon (14.42), Adj (17.59), and M/M/P (11.34). For the Bon category, the observed values are again almost opposite of what is expected, with idiosyncrasies tending not to involve the functor. The Adj category also follows this same pattern: the tendency is for the Func--Y cell to account for fewer than the expected number of idiosyncratic collocations, and the deviance values are even higher here than in the Bon category. M/M/P contributes measurably to the total deviance, but for a different reason. This category experiences a greater than expected tendency for the idiosyncrasy to involve the functor.

A qualitative analysis of the data in the Bon, Adj, and M/M/P categories is necessary at this point to help interpret the high deviance values.

5.3.3 Qualitative analysis of headword and functor

The categories contributing the most to the deviances in tables 48 and 49 were Bon, Adj, and M/M/P. A closer look at the actual collocations in these categories may explain why they returned such divergent results. Table 51 shows the idiosyncratic collocations included in the Bon category, separated according to whether the idiosyncrasy is contained in the headword (H), functor (F), or both (H/F). Idiosyncratic words are printed in bold type.

TABLE 51

BON CATEGORY WITH HEADWORD/FUNCTOR DISTINCTION

ST	Student Phrase	Target Collocation	Part
GO	ethic in value concepts	ethical values	H
GO	ethic value concepts	ethical values	H
CA (2)	value concepts	value judgments	H
WI (2)	value concepts	value judgments	H
DI (4)	value concepts	value judgments	H
WI	rooted concepts	deeply-rooted convictions	H
CA	luxury objects	luxury items	H
CA	expensive objects	luxury items	H/F
CA (2)	original clients	loyal customers	H/F
LY (2)	high class cloth	expensive clothing	H/F
CA (2)	unique flavor	distinctive taste	H/F
DI	definite customers	loyal customers	F
LY	faithful on the same products	loyal to the same products	F

Five of the eight students contributed to the Bon category. The headword of all the collocations is a noun and the functor is an adjective. The striking feature is that eleven of the 21 collocations in this category involve the use of the word *concepts* as an idiosyncratic replacement for the words *values*, *judgments*, and *convictions*. The FolioVIEWS database confirms that aside from CA and WI (who wrote their collocations in the same two assignments), these collocations were written in different assignments on different dates. Even if these collocations were not included, the Bon category still heavily favors

idiosyncratic nouns in adjective + noun collocations. Perhaps this is due to the fact that many of the headwords and all of the functors in this category are very specific to the business genre, while other categories contain collocations which are more widely used in general English contexts.

The Adj category is also a candidate for scrutiny, having a large number of idiosyncratic headwords. Table 52 shows the idiosyncratic collocations included in the Adj category, separated according to whether the idiosyncrasy is contained in the headword (H), functor (F), or both (H/F). Idiosyncratic words are printed in bold type.

TABLE 52

ADJ CATEGORY WITH HEADWORD/FUNCTOR DISTINCTION

ST	Student Phrase	Target Collocation	Type
LU (2)	business portion	business segment	H
GO (3)	business class	business market/segment	H
GO (2)	business class to student class	business segment ...	H
CH (2)	better health technique	better health habits	H
PO	complex works	complex tasks	H
LU	market characteristic	market trends	H
PO	the different education style	a different education system	H
DI (3)	the product figure	the product profile	H
LU (2)	pay practice	wage/pay scale	H
CH	a certain years	a certain age	H
DI (2)	purchasing abilities	purchasing power	H
CH	the average life	the average age	H
PO	to use their political rights	to use their political power	H
LU	living space	living conditions	H
GO (2)	working time	hours of operation	H/F
PO	the salary notice	the pay slip	H/F
WI	consuming ability	consumption potential	H/F
CH	the aging era	old age	H/F
PO	middle level persons	middle-class people	H/F
GO	egalitarian society to money society	capitalistic society	F
CA (4)	initial customers	new customers	F
LY	portion for marketing	aspect of marketing	F
CH (4)	market share rate	market share	F
LY	raise the crimes committing rate	raise the crime rate	F
CH	both physical and mind condition	mental condition	F
PO (3)	customer psychology	consumer psychology	F
LY	in this rival society	in this competitive society	F
LU	unbenefit behavior	dangerous behavior	F
GO (2)	attendance cards	time cards	F
PO (2)	the routine law service	routine legal services	F
CH	years of their rest life	years of their retired life	F

All eight students contribute to the Adj category. Just as in the Bon category, these are all adjective + noun combinations in which the noun is the headword. Unlike the Bon category, students had problems with these nouns as often as they had problems with the adjectives. The overall pattern in the data is for the idiosyncrasy to occur in the functor 70% of the time. One possible contributor to the 50-50 spread between functor and headword in the Adj category is the occurrence of collocations that would not be labeled learner

collocations in another context. That is, of the 23 cases in which a collocation involves idiosyncrasy in only the headword, 3 (accounting for 9 total occurrences) are NS collocations in other contexts: business class (when talking about airplane seating), the product figure (when talking about profits and/or losses), and political rights (when talking about repressed groups).

The M/M/P group is actually composed of Magn, Mult, and Plus. The reason this group had a significant chi-square value is because of its overall scarcity of headword idiosyncrasy. Each of the categories in the group is first analyzed separately and then discussed as a whole.

Table 53 shows the idiosyncratic collocations included in the Mult category, separated according to whether the idiosyncrasy is contained in the headword (H), functor (F), or both (H/F). Idiosyncratic words are printed in bold type.

TABLE 53

MULT CATEGORY WITH HEADWORD/FUNCTOR DISTINCTION

ST	Student Phrase	Target Collocation	Type
CH	increasing number of older generation	increasing size	F
LU	a number of evidence	a wealth of evidence	F
WI	a great number of money	a great deal of money	F
GO	a number of money	a lot of money	F
CH (3)	our middle life	our middle-age years	H/F
DI (2)	a large part of the people	a large number of people	F
LY	grand amount of currency exchange	large amount	F
LY	large amount of dollars	large amount of money	H
LY	spend so many time	spend so much time	F
CH	the rate of the consumers	the number of consumers	F
CH	they still own numerous knowledge	vast knowledge	F
PO (2)	keep the whole assets	keep all the assets	F
CH (3)	abundant working experience	a wealth of experience	F
LU	in the rest year	in the other years	F
LU (2)	the portion of female user	percentage of female users	F
LU	requirements including housing , clothing and eating	necessities including food, clothing and shelter	H/F

Seven of eight students contributed to the Mult category, which consists of 23 noun phrase collocations. Eleven of the idiosyncrasies appeared in the functor of a collocation in which a count versus non-count noun distinction was required: each of the seven students contributed at least one of these. Since Mandarin does not distinguish between count and non-count nouns, it is possible that native language interference contributed to the unusually high number of learner collocations involving functor idiosyncrasy in this category.

Table 54 shows the idiosyncratic collocations included in the Plus category, separated according to whether the idiosyncrasy is contained in the headword (H), functor (F), or both (H/F). Idiosyncratic words are printed in bold type.

TABLE 54

PLUS CATEGORY WITH HEADWORD/FUNCTOR DISTINCTION

ST	Student Phrase	Target Collocation	Type
CA (3)	raise the number of advertisements	increase the number	F
CA	raising the number of advertisements	increasing the number	F
DI	Polio can raise its sales	Polio can increase its sales	F
DI	rise its sales	increase its sales	F
CA	attain many edges	gain an edge	F
LU	gain more competitive	become more competitive	F
GO	heighten his quality of life	improve his quality of life	F
CA	reinvigorate the market share	increase the market share	F
CA	retrieve old customers	regain former customers	F
GO (2)	suffered from ... improvement	benefited from	F
CH	take a lot of advantages by	gain a lot of advantages by	F
CH	to grasp knowledge	to acquire knowledge	F
LY	build up their own features	build up ... reputations	H
LY	the backup of the greenback	reemergence of ...	F
LU	to gain more competition	to gain a larger share ...	H
LY	filled with explosive technology	increasing technology	F
PO	by learning knowledge	by acquiring knowledge	F
WI	further realize	further understand	H
LY	push the US dollar's rising	push the US dollar upward	F

Of the 22 collocations produced by seven students in this category, 21 contain verbs and 18 are verb + noun pairs in which the noun is the headword. Sixteen of those 18 contain an idiosyncrasy in the verb. Two students are responsible for six collocations in which the verbs *rise* and *raise* are used in place of *increase*, and another uses the adjective *explosive* in place of *increasing*. In addition, the word *gain* was the cause of functor idiosyncrasy in three collocations. Part of the reason for the unusual behavior of this category is the use of these words. Eliminating these collocations from this category results in a rate of 75% functor idiosyncrasy, which is in line with the expected 70% functor idiosyncrasy.

Table 55 shows the idiosyncratic collocations included in the Magn category, separated according to whether the idiosyncrasy is contained in the headword (H), functor (F), or both (H/F). Idiosyncratic words are printed in bold type.

TABLE 55

MAGN CATEGORY WITH HEADWORD/FUNCTOR DISTINCTION

ST	Student Phrase	Target Collocation	Type
WI	the more possible	the greater the possibility	F
GO (2)	to pay more salaries to employees	to pay higher salaries	F
LU (2)	pay more salary	pay higher salaries	F
LU (3)	more market share	larger market share	F
DI	huge different action	largely different action	F
PO	huge abilities	great abilities	F
GO	so huge abilities	such great advantages	H/F
CA (2)	deeper status within each market	better/higher status	F
CH	their too cautious	their extreme caution	F
WI	in the unknown future	in the distant future	F
LY	long options for hedging	long-term options	F
DI	the aging generation	the older generation	F
CH	the expedition of the human life	increased life expectancy	H/F
CA	the higher desires	higher-level desires	F
DI	to lack well support	to lack good/strong support	F
DI	an intense impression	a strong impression	F
CA	forceful evidence	strong evidence	F
WI	extremely possible	extremely likely	H
WI (2)	greatly possible	extremely likely	H/F
GO (2)	fully important	extremely important	F
WI	enlarge the distance	increase the distance	F
WI	widen his knowledge	broaden his knowledge	F
WI (3)	accepted broadly	widely accepted	F

The Magn category contains 22 adjective + noun collocations. In all of these cases, the adjective is the site of an idiosyncrasy. This is in stark contrast to the Bon and Adj categories, in which nouns are often the site of the idiosyncrasy. In addition to the adjective + noun collocations, there are five adjective + adjective collocations, all involving the word *extremely*. There are two noun + verb pairs in which the idiosyncrasy lies in the verb, and three adverb + verb pairs in which the adverb is the site of the idiosyncrasy. An overgeneralization of the word *more* contributed to eight of the 22 adjective + noun collocations, and there were four cases in which the word *extremely* was missing from the functor of a collocation. Removing these 12 cases from this category does not result in a significant drop in the percentage of functor idiosyncrasies.

The three categories in the M/M/P group show a strong tendency not to have idiosyncrasy in the headword, which is most often a noun. Only fourteen headword nouns were idiosyncratic in the 77-member set of collocations in this group. Students clearly are capable of producing the headwords in these categories. Lack of collocational knowledge of particular words (i.e., increase, gain, more, and extremely) contributed measurably to the high deviance values for this group. In addition, the lack of a count versus non-count noun distinction in Mandarin may have contributed to the deviance value.

5.3.4 Examination of shared idiosyncrasies

The results of the functor versus headword analysis of the learner collocations shows that the site of the idiosyncrasy was most commonly either the headword or the functor, but rarely both. This section will examine the 30 collocations that involved idiosyncrasy at both sites. Table 56 lists all the collocations with both headword and functor idiosyncrasies.

TABLE 56

COLLOCATIONS INVOLVING BOTH HEADWORD AND FUNCTOR

IDIOSYNCRASY

ST	Student Phrase	Target Collocation
CH (3)	our middle life	our middle-age years
PO	middle level persons	middle-class people
GO	so huge abilities	such great advantages
CH	the expedition of the human life	increased life expectancy
CA (2)	original clients	loyal customers
LY (2)	high class cloth	expensive clothing
CA (2)	unique flavor	distinctive taste
CA	expensive objects	luxury items
CA	impending era	coming years
GO (2)	working time	hours of operation
PO	the salary notice	the pay slip
CH	the aging era	old age
WI	consuming ability	consumption potential
LU	requirements including	necessities including
	housing, clothing and eating	food, clothing and shelter
GO	assign time recording	enforce time-keeping
CH	declining the rate of sick	reducing the rate of illness
PO	contact the topic of inheritance	encounter the issue of inheritance
CH	to replace their jobs	to fill open positions
LU	realize how the market is	discover the condition of the market
LU (2)	to prove my guess	to confirm my suspicions
PO	fit the desires of them	meet their needs
WI (2)	greatly possible	extremely likely

Learner collocations involving both functor and headword idiosyncrasy were produced by seven of the eight students and appear in nine of the 20 Mel'čuk categories. Twenty are noun phrase collocations, eight are verb + noun collocations, and two are adjective + adjective collocations. This particular perspective of these collocations does not appear to reveal any particular patterns setting these learner collocations apart from those in which only the functor or the headword is idiosyncratic.

5.4 A proposed characterization of learner collocations

Chapters 4 and 5 have presented the application of a combination of quantitative and qualitative analyses to learner collocations from the corpus of student writing. The results of these analyses reveal several interesting facts about learner collocations; the results of each step in the analyses are discussed below. Taken together, these results lead to a tentative characterization of learner collocations.

In the initial qualitative analysis, learner collocations were divided into four types, according to greatest probable influence: lexical transfer, phonology, grammar, and semantics. Of these four types, those involving lexical transfer account for 12% of the total; those involving phonology account for 12.5%; those involving grammar account for 23.5%; and those involving semantics account for 52%. Thus, the first characteristic of learner collocations is that a large number of them involve semantic idiosyncrasy as their most salient feature.

Once the learner collocations were divided up into four types, a quantitative analysis (the chi-square test) of all possible influences on the idiosyncrasies showed that those collocations involving an idiosyncrasy in form (the phonology and grammar categories) tended not to also involve an idiosyncrasy in meaning (the lexical transfer and semantics categories), and vice versa. Thus, the second characteristic of learner collocations is that they tend to involve idiosyncrasy in either form or meaning, but rarely both.

Further qualitative analysis of those collocations involving a form (grammatical) idiosyncrasy revealed that it is possible students are relying in some cases on a native language structure. This observation was tested quantitatively with a chi-square test on two possible categories of the source of the idiosyncrasy: reliance on a native language structure, or reliance on another, unknown source. The results of this test provide the third characteristic of learner collocations: those involving an idiosyncrasy in preposition tend to

imitate native structure, while those involving an idiosyncrasy in word form must always imitate another source.

Further qualitative analysis of those learner collocations involving a meaning (semantic) idiosyncrasy was achieved by splitting each collocation into two parts: a headword and a functor. A count of headword and functor idiosyncrasies revealed that the idiosyncrasy is more often present in the functor than the headword and rarely occurs in both; of the one category (Bon) in which a headword idiosyncrasy is more common, the high number of headword idiosyncrasies is due to a few words. Thus, the fourth characteristic of learner collocation is that the site of the idiosyncrasy is most likely the functor and rarely occurs on both parts of the same collocation. For those collocations involving an idiosyncrasy in the headword, the cause is likely one of a restricted group of words.

The implications of these characteristics will be explored in chapter 6.

CHAPTER 6

CONCLUSIONS

6.0 Summary and conclusions

This study has investigated learner collocations in the English writing of eight speakers of Mandarin who attended a special business English class at The University of Texas at Arlington during the Fall 1995 semester. All assignments written by each of the students were collected and scrutinized for the occurrence of learner collocations. The collocations themselves were then divided into groups and examined in a series of qualitative and quantitative analyses. The objective of the study was to synthesize from these analyses a characterization of learner collocation. Results revealed specific patterns in learner collocation, which led to the development of such a characterization.

An initial qualitative analysis led to the identification of four categories of idiosyncrasy in learner collocation: lexical transfer, and phonological, grammatical, and semantic idiosyncrasy. Those collocations containing a semantic idiosyncrasy comprised more than half the total. Each of the four categories was then further subdivided according to the exact nature of the idiosyncrasy. A chi-square test run for two-by-two contingency tables for the four basic categories indicated a strong tendency for the collocations to contain an idiosyncrasy in either form or meaning, but rarely both; chi-square tests for each individual student's data affirmed this finding with very similar results. Since the semantic and grammatical categories showed the most evidence of opposition in the chi-square test, each was selected for further analysis. For learner collocations involving a grammatical idiosyncrasy there were three important findings: when the preposition is the site of the

idiosyncrasy, the learner tends to resort to native-speaker language structures; when the form of a word is the site of the idiosyncrasy, the learner tends to resort to something other than native-language structures; when the site of the idiosyncrasy is another structural element, such as a genitive structure, the learner tends to resort to either native-speaker language structures or something other than native-language structures, on a random basis. Learner collocations involving a semantic idiosyncrasy were further analyzed into headword and functor parts, of which the functor was most often the site of the idiosyncrasy. A chi-square test was run on a set of collapsed subdivisions of the semantics category, which showed that for certain of the subdivisions the site of the idiosyncrasy was more likely to be a headword than a functor. This was at least partially accounted for by multiple occurrences of the same idiosyncratic word.

6.1 Implications

One of the most important outcomes of this study is that the researcher has developed a primary classification system for learner collocations based on a corpus of learner collocations in context. In most published studies, learner collocations are typically categorized according to native speaker collocation categories, such as that proposed by Benson et al. (1986); instead of predetermining the categories, the researcher in this study has determined the categories based on full-text data produced by learners. The identification of four categories of idiosyncrasy (lexical transfer, phonological, grammatical, and semantic) therefore fill a void in the current literature on learner collocations. In addition, learner collocations have heretofore been studied almost exclusively with recognition tasks in which specific collocations are elicited (e.g., cloze exercises in which one part of the collocation is omitted). The data collected here are subject-directed, not researcher-directed. The end result is that the categories uncovered

here can be used in replication studies for native speakers of other languages, or for native speakers of Mandarin in other specialized language domains.

A second implication involves the set of collocation categories themselves. Especially important is the finding that overall, these learner collocations involved an idiosyncrasy in either form or meaning, but seldom both. This suggests that when the learner produces language, he or she may be able to access the lexicon through multiple pathways; specifically, the results of this study indicate that form and meaning aspects of the lexicon may occupy separate pathways. The finding that collocations in the semantic category tend to involve idiosyncrasy in the functor more often than in the headword seems to indicate that learners are more familiar with the exact shade of meaning carried by semantic headwords, which would mean that what learners need the most help with are functors. This has further implications for ESL teachers, who presently have no learner collocation-based analysis in which to situate their approaches to teaching collocations in the classroom. Current approaches to determining which collocations need to be taught tend to rely on the frequency with which native speaker collocations occur in texts. The results of this study provide teachers with some information about the types of collocations students are most likely to have trouble producing.

The breakdown of learner collocations into categories also provides insight into the nature of the native speaker collocation. Studies of native speaker collocations of course do not need to consider the issue of lexical transfer, or phonological idiosyncrasy. What could be gained through a comparison of NS categories and the categories proposed in this study of learner collocations is a more thorough understanding of collocation in general. This is important information which provides a new perspective on the phenomenon that should inform approaches to the analysis of NS collocation.

This study has also shown how computer tools can be successfully applied to a learner corpus to search for answers to the research question motivating this study. Various

computer programs were used in both the qualitative and quantitative analyses to provide a corpus-driven characterization of learner collocations. In particular, this study has demonstrated the value of a program such as FolioVIEWS, with which the researcher can systematically examine very specific aspects of a corpus, yet always have the original context to return to.

On the issue of whether native language influences idiosyncratic collocation, the results of this study indicate that native language decidedly influences learner collocations. The first category identified in this study is the influence of the native language on learner collocations. Past research often cites transfer as a cause in the construction of idiosyncratic collocation, but this study takes that further by identifying at least three types of transfer affecting collocations produced by native speakers of Mandarin: direct lexical translation (i.e., **open the computer*); lexical overgeneralization due to a mismatch between English and Mandarin (i.e., *make* and *do*); and structural item transfer (i.e., prepositions and genitive structures). The fact that this study found three types of transfer suggests that there may be other types. In addition, the fact that idiosyncratic collocation due to lexical transfer comprises the smallest group in this study indicates that earlier studies may not have gone far enough in their analyses.

The second category of learner collocation discussed in this study involves phonological error. This influence on learner collocation has been rarely mentioned; it is possible that what is labeled phonological influence in this study has been classified as native language influence in other studies. For this particular group of learners the end of a word was the most likely site for an idiosyncrasy, which complements the fact that Mandarin is for the most part not an inflected language. In other words, native speakers of Mandarin do not need to pay special attention to the ends of words. Regardless of the label, the learner collocations in the phonological category suggest that sound perception is a factor in at least the reception of collocations. None of the learners in this study ever read

**safe heaven, *result the problem, or *put in a blind.* But perhaps they think they heard or read these phrases. In other words, in accessing a word in the lexicon, students activate like-sounding lexical items, and perhaps remember the more familiar-sounding word. It is also possible that during production the student builds the collocation word-by-word, and at that point inserts a like-sounding word in place of what he or she originally heard. So, while the phonological idiosyncrasies might be attributable to native language influences, they may also involve issues of accessing the lexicon. The existence of a significant number of these idiosyncrasies suggests that a systematic review of the phonology of English, especially in regard to word endings as distinguishing features, might be more helpful to learners than a systematic review of the grammar of English inflection.

Idiosyncratic collocation due to grammar was the third category of learner collocation identified in this study. Word form, preposition choice, and genitive structures proved to be the main problems. Word form may be a problem for native speakers of Chinese because Mandarin has no equivalent phenomenon. For example, an adjective can become a noun but it does not change its form. In some ways this is similar to the phonology problem with word endings: the learner must switch from an unmarked system (Chinese morphology, where words change little if at all) to a marked system (English morphology, where word endings and word forms change often). This suggests that lessons drawing attention to the importance of word form may help native speakers of Mandarin to produce more structurally accurate collocations. Genitive structure problems such as **live's quality* and **styles of work* indicate a fundamental difference in modification between Mandarin and English—Mandarin has one genitive structure while English has two. Again, the student must move from a less complex system to a more complex system.

The category with the most learner collocations was semantic: these occurred consistently throughout the assigned writings with greater frequency than any of the other

collocation types. This is significant, because it points to the area in the acquisition of collocations that requires the greatest attention. And the fact that the learner collocations in this category involved 20 different lexical relations shows that learners are struggling consistently with many different issues, not just one or two. In addition, this is the group of collocations which most clearly reveal whether or not a learner has a knowledge of appropriate word combination, since the student is replacing the appropriate word(s) with what he or she thinks is a synonym(s). So, for example, learners produced **work out poverty*, **prove my guess*, and **luxury objects* without realizing these phrases do not represent standard written American English. If it is true that this category encompasses those collocations most often written idiosyncratically, then performance in this category alone may prove to be an adequate indicator of a student's proficiency in a given language domain.

In addition to the implications of the categories of learner collocations identified in this study are the implications of how many times certain error occurred. For example, none of the students were able to present correct forms of collocations involving *make* or *do* on subsequent assignments. Only one student, CA, presented correct and incorrect forms of the same collocation, but they were in the same assignments. CA produced two instances of *make a survey* in the second and third drafts of the marketing paper, while at separate points in those same papers producing two instances of *do a survey*. This has interesting implications for those studies in which students are tested for a single instance of a collocation. In other words, just because a student produced a collocation correctly once does not mean that he or she "knows" a collocation. The evidence presented here suggests that students who do produce correct collocations on one occasion can not necessarily be said to have learned them.

The results of this study show that a characterization of learner collocations opens up new considerations for research on learner language and the creation of materials for the

NNS. A characterization of learner collocation, for example, should be an important component of both teaching and reference material aimed at the NNS. Frequency of error information alone should be a consideration in the design of a collocational dictionary. In addition, the results of this study can now be tested against similar results for native speakers of other languages, and in other contexts. Further studies of English collocation with native speakers of other languages could, together with results such as those presented by this study, identify universal characteristics of learner collocations, especially in relation to language transfer.

6.2 Directions for further research

This study puts forward a tentative characterization of learner collocations. The results suggest that the division of the collocations into four categories based on the general nature of their idiosyncrasy is a reasonable one. Further breakdown of those categories into subtypes, however, may be less valid. In fact, it is possible that the subtypes of the semantic category in particular have little value in the characterization of learner collocations. More work is needed to confirm these findings. Furthermore, a test of the primary classification scheme on native speakers of other languages is needed in order to confirm whether it is possibly a universal classification, or only language-specific.

The reading corpus allowed for checks of native speaker versions of the idiosyncratic collocations. Unfortunately, the reading corpus for this study was so small that native speaker versions of collocations pertinent to this study were rare; no substantial conclusions could be drawn from these results. Future studies should take into account this weakness; this aspect of the study may have been more useful if all reading materials used by each student in the creation of the research paper had been included in the reading corpus. An accurate measure of exposure to collocations is necessary information in any investigation of collocation acquisition.

One other possible direction for further research involving word frequency is the comparison of idiosyncratic words in learner collocations with a basic English word frequency list. Are the words learners have difficulty producing specific to the language domain within which they are writing, or do their problems involve basic English vocabulary? The results of such a study may increase understanding of the nature of idiosyncratic collocation.

In terms of characterizing learner collocations, a basic form versus meaning opposition has been uncovered in the course of this analysis of learner collocations. Further study is needed to replicate this finding, both to affirm the validity of the opposition and to add evidence to the multiple pathway theory of lexical access. This in turn could be followed up with an analysis of a spoken corpus: How are the results of an analysis of such a data set different from those observed here?

Although this study did not expressly address the issue of collocation in learner dictionaries, one of the results of this study is a compilation of collocations that NNS students had trouble producing. It is this type of information, on a larger scale, that compilers of learner dictionaries need when deciding what sorts of collocational information to include in reference materials. Dictionaries for learners of English such as the *Oxford Dictionary of Business English* (1993) traditionally use native speaker-produced texts and native speaker intuition in making decisions about what collocational information should be included in individual entries. Further studies of learner collocation in which large numbers of idiosyncratic collocations are collected from native speakers of various native languages would allow compilers of dictionaries that are aimed at the NNS to consider including in their materials at least some collocational information for which there is a consistently demonstrated need.

The results of this study affect an understanding of collocation in general by problematizing the definition itself, and looking at collocation as a function of what the

NNS has trouble producing. This is an area which needs much more study, but more in-depth studies of the nature of learner collocations could eventually lead to a more uniform definition of collocation in the applied linguistics literature.

As for the problem of teaching collocations to NNSs, further studies are needed so that ESL instructors are informed enough about learner collocations to be able to devise appropriate materials to effectively teach them. Traditionally, materials for teaching collocations in the ESL classroom have been based on frequency, if that information is available, or the intuition of the teacher about what will be most useful. The results of this study suggest that it is possible to base decisions about what to teach on a characterization of actual observed problems. The American business collocations used in materials by Flower and Martínez (1995), for example, could be compared with the idiosyncratic collocations found in this study. Exercises containing collocations students had trouble with could then be used in the classroom in a focused lesson.

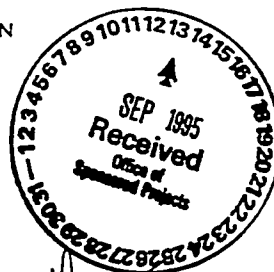
Another aspect of teaching collocations not covered in this study involves the addition of an experimental group taught about collocation at the beginning of the semester and coached in their idiosyncrasies. One would expect that such a group would produce fewer idiosyncratic collocations. This type of study could help identify those learner collocations, or types of learner collocations, that could most easily be taught. It is that sort of information—which learner collocations can be improved—that would be of most help to those who are involved in the compilation of teaching materials, teacher references, and collocational reference materials.

APPENDIX A

HUMAN SUBJECTS COMMITTEE APPROVAL LETTER
AND CONSENT FORM



A Tradition of Excellence. A Future of Opportunity.
THE UNIVERSITY OF TEXAS AT ARLINGTON



MEMORANDUM

DATE: September 15, 1995

TO: Office of Sponsored Projects

FROM: Karen B. Heusinkveld, Dr. PH
Chairman of Human Research Review Committee

SUBJECT: Principal Investigator: Robin Lombard/ Humanities, Linguistics

Title of Prospectus: Work Combination in International Student Writing

The Human Research Review Committee has concluded that the above named prospectus poses no risk to human subjects and qualifies for exempt status under current Department of Health and Human Services regulations (45 CFR. Part 46). No further review of the project is therefore required. However, it should be noted that any change in the study as it is described in the prospectus could affect its status and necessitate HRRC review.

SCHOOL OF NURSING

Informed Consent

I understand that the study I will participate in this semester involves the collection of all written work I will do for this class. This written work will be placed in a computer database by the researcher, and later analyzed for vocabulary use. In addition, I will be asked to fill out a questionnaire asking about my language background and abilities in English. At any time during the semester I will have the opportunity to ask questions regarding the study. A full explanation of the study will be given to me at the end of the semester.

No discomfort is anticipated, since the study will not affect the normal assignments and procedures of the class.

The major benefits I will receive from participation in this study is the knowledge that I contributed to the greater understanding of how International students use written English.

I understand that participation in the study is not a requirement of this class, and will have no effect on normal classroom policies and procedures.

I understand that records of my participation in this study will be held strictly confidential. My identity as a subject will not be disclosed to anyone beyond the researcher.

This research is under the supervision of Robin Lombard, MA, ABD. Ms. Lombard's office is 330 Hammond Hall and her phone number is (817) 273-3052. Please feel free to contact Ms. Lombard if you have any questions.

I hereby consent to participate in this study
and understand the above explanation.

Signature

Date

APPENDIX B

STUDENT QUESTIONNAIRES

CA

Age: 26

Sex: Male

What is your home country? Taiwan

What is your native language? Mandarin

Do you speak/read/write any other languages besides your native language and English?

No

If so, what languages? _____

How long did you study English in your home country? 7 years

How long have you been in the US? 3 months

Did you study in a special English language program in the US before you came to UTA?

No

If so, where? _____

What degree are you studying for at UTA? Master's in Real Estate

Have you studied this subject in your native language? No

If so, for how long? _____

What was your best TOEFL score? 550

Did you take GMAT or GRE? GMAT

What was your best score? 520

What was your undergraduate major? Public Finance and Taxation

What graduate classes are you taking this semester? Quantitative Analysis

How many hours of reading in English do you think you do every week (homework plus any other reading you might do)? about 20 hours

CH

Age: 23

Sex: Female

What is your home country? Taiwan

What is your native language? Mandarin

Do you speak/read/write any other languages besides your native language and English?

No

If so, what languages? _____

How long did you study English in your home country? 8 years

How long have you been in the US? 3 months

Did you study in a special English language program in the US before you came to UTA?

No

If so, where? _____

What degree are you studying for at UTA? Master's in Business Administration

Have you studied this subject in your native language? Yes

If so, for how long? 4 years

What was your best TOEFL score? 550

Did you take GMAT or GRE? GMAT

What was your best score? 560

What was your undergraduate major? Accounting

What graduate classes are you taking this semester? None (MBA Bridge Program)

How many hours of reading in English do you think you do every week (homework plus any other reading you might do)? 20 hours

DI

Age: 22

Sex: Female

What is your home country? Taiwan

What is your native language? Chinese (Mandarin)

Do you speak/read/write any other languages besides your native language and English?

Yes

If so, what languages? Japanese

How long did you study English in your home country? 10 years

How long have you been in the US? 3 months

Did you study in a special English language program in the US before you came to UTA?

No

If so, where? _____

What degree are you studying for at UTA? Master's in Business Administration

Have you studied this subject in your native language? Yes

If so, for how long? 4 years

What was your best TOEFL score? 560

Did you take GMAT or GRE? GMAT

What was your best score? 550

What was your undergraduate major? Accounting

What graduate classes are you taking this semester? Marketing, Introduction to Management Science

How many hours of reading in English do you think you do every week (homework plus any other reading you might do)? 40 hours

GO

Age: 25

Sex: Male

What is your home country? Taiwan

What is your native language? Chinese (Mandarin)

Do you speak/read/write any other languages besides your native language and English?

No

If so, what languages? _____

How long did you study English in your home country? 8 years

How long have you been in the US? 5 months

Did you study in a special English language program in the US before you came to UTA?

No

If so, where? _____

What degree are you studying for at UTA? Master's in Public Administration

Have you studied this subject in your native language? Yes

If so, for how long? 3 months

What was your best TOEFL score? 510

Did you take GMAT or GRE? GMAT

What was your best score? 520

What was your undergraduate major? Cooperative Economics

What graduate classes are you taking this semester? None (MBA Bridge Program)

How many hours of reading in English do you think you do every week (homework plus any other reading you might do)? 10 hours

LU

Age: 26

Sex: Male

What is your home country? Taiwan

What is your native language? Chinese (Mandarin)

Do you speak/read/write any other languages besides your native language and English?

No

If so, what languages? _____

How long did you study English in your home country? 10+ years

How long have you been in the US? 2 months

Did you study in a special English language program in the US before you came to UTA?

No

If so, where? _____

What degree are you studying for at UTA? Master's

Have you studied this subject in your native language? Yes

If so, for how long? 4 years

What was your best TOEFL score? 540

Did you take GMAT or GRE? GMAT

What was your best score? 560

What was your undergraduate major? Statistics

What graduate classes are you taking this semester? Management, Quantitative Analysis

How many hours of reading in English do you think you do every week (homework plus any other reading you might do)? 48 hours

LY

Age: 27

Sex: Female

What is your home country? Taiwan

What is your native language? Mandarin

Do you speak/read/write any other languages besides your native language and English?

No

If so, what languages? _____

How long did you study English in your home country? 10 years

How long have you been in the US? 3 months

Did you study in a special English language program in the US before you came to UTA?

No

If so, where? _____

What degree are you studying for at UTA? Master's

Have you studied this subject in your native language? No

If so, for how long? _____

What was your best TOEFL score? 540

Did you take GMAT or GRE? GMAT

What was your best score? 550

What was your undergraduate major? Public Finance

What graduate classes are you taking this semester? Management, Statistics

How many hours of reading in English do you think you do every week (homework plus any other reading you might do)? 30 hours

PO

Age: 28

Sex: Female

What is your home country? Taiwan

What is your native language? Mandarin

Do you speak/read/write any other languages besides your native language and English?

Yes

If so, what languages? Taiwanese, Japanese

How long did you study English in your home country? 10 years

How long have you been in the US? 3 months

Did you study in a special English language program in the US before you came to UTA?

No

If so, where? _____

What degree are you studying for at UTA? Master's in Business Administration

Have you studied this subject in your native language? Yes

If so, for how long? 4 years

What was your best TOEFL score? 510

Did you take GMAT or GRE? GMAT

What was your best score? 490

What was your undergraduate major? Business Administration

What graduate classes are you taking this semester? None (MBA Bridge Program)

How many hours of reading in English do you think you do every week (homework plus any other reading you might do)? 15 hours

WI

Age: 25

Sex: Female

What is your home country? Taiwan

What is your native language? Chinese (Mandarin)

Do you speak/read/write any other languages besides your native language and English?

No

If so, what languages? _____

How long did you study English in your home country? 13 years

How long have you been in the US? 2 months

Did you study in a special English language program in the US before you came to UTA?

No

If so, where? _____

What degree are you studying for at UTA? Master's

Have you studied this subject in your native language? No

If so, for how long? _____

What was your best TOEFL score? 553

Did you take GMAT or GRE? GMAT

What was your best score? 540

What was your undergraduate major? Information Systems

What graduate classes are you taking this semester? Advanced Statistics, Database (INSY 5335), Networks (INSY 5343)

How many hours of reading in English do you think you do every week (homework plus any other reading you might do)? 10 hours

APPENDIX C

READING TEXT BIBLIOGRAPHY

Reading Text Bibliography

- Chao, Julie. 1995. China turns to Confucius for help with ethical crisis. *Dallas Morning News*, Feb. 16: 37A.
- Elmer-Dewitt, Philip. 1992. Preparing for the worst. *Exploring through Writing*, 2nd edition by Ann Raimés, 180-183. New York: St. Martin's Press.
- Goodman, Ellen. 1987. Don't call me, I'll call you - but not from cloud 9. *Reading on purpose* by Fraida Dublin and Elite Olshstein, 111-112. Reading, MA: Addison-Wesley.
- Ho, Kie. 1987. We should cherish our children's freedom to think. *Reading on purpose* by Fraida Dublin and Elite Olshstein, 132-133. Reading, MA: Addison-Wesley.
- Johnson, Robert. 1995. A problem solving case: marketing. Unpublished manuscript.
- Kolter, Philip. 1988. Market targeting. *Academic reading: A content-based approach* ed. Louis W. Holschuh and J. Patrick Kelley, 67-70. New York: St Martin's Press.
- Kolter, Philip. 1992. *Reading on marketing. Focus on business* by Keith Maurice, 56-57. Englewood Cliffs, NJ: Regents/Prentice Hall.
- Kübler-Ross, Elisabeth. 1987. The fear of dying. *Reading on purpose* by Fraida Dublin and Elite Olshstein, 156-159. Reading, MA: Addison-Wesley.
- Kurtz, David L. and Louis E. Boone. 1988. Consumer behavior: Basic concepts. *Academic reading: A content-based approach*, ed. Louis W. Holschuh and J. Patrick Kelley, 25-35. New York: St Martin's Press.
- Leo, John. 1991. A welcome sign for immigrants. *U.S. News and World Report*. September 16: 18.
- Williams, Walter. 1994. How to really help people escape poverty. *The Dallas Morning News*, July 2: 31A.
- Wilson, Laurie. 1995. Job skills or knowledge? Students differ on college goals. *Dallas Morning News*, Nov. 28: 22A.

APPENDIX D

COLLOCATION CROSS-CLASSIFICATION
FOR EACH STUDENT

TABLE

Collocation Cross-Classification for Each Student -- CA

Collocation	Lex Transfer	Phon	Gram	Sem
a characteristic of distinctive			X	
a trend of change			X	
act a role like				X
attain many edges				X
bond to think		X		X
borrow urgent funds (2)	X			
can't successfully work out poverty				X
carry lots of contributions				X
change in strategy should be done	X			
content the basic needs			X	X
deeper status within each market (2)				X
expend the market share		X		
expensive objects				X
express specialty (2)				X
express their distinctive flavors				X
forceful evidence				X
four-fifth of all single-family home loans (2)		X	X	
full satisfaction (2)				X
fully realize (2)				X
impending era				X
in order to make references		X		X
initial customers (4)				X
keep the serving of the loan		X		X
keep their eyes for			X	X
lower loan-to-value ratios		X		
luxury objects				X
make a marketing survey	X			
make a wrong definition	X			X
make marketing survey	X			
make some of the analysis	X			
making another survey	X			
Mostly important			X	
original clients (2)				X
over third-fourth (2)	X	X		X
price of food will raise		X		X
proceed a market poll				X
provide the loan serving		X		X
put in a blind		X		
raise the number of advertisements (3)				X
raising the number of advertisements				X
realize the reason				X
reinvigorate the market share				X
relatively investing concepts (2)		X		X
rescue the loss (4)		X		X
retrieve old customers				X
sales are still monotonous				X

TABLE (Continued)

Collocation Cross-Classification for Each Student -- CA				
Collocation	Lex Transfer	Phon	Gram	Sem
satisfaction of necessity				X
securities backed on the mortgages (2)			X	
support this standpoint		X		X
the class of business and working (2)			X	
the higher desires				X
the seriousness of a jeopardy			X	
to reserve this drawback		X		
twice times	X			X
unique flavor (2)				X
value concepts (2)				X
what inclinations do most students prefer				X
which roles do we want to act				X

TABLE

Collocation Cross-Classification for Each Student -- PO

Collocation	Lex Transfer	Phon	Gram	Sem
a lower tax rate basket (2)		X		X
apply work			X	
As can been seen (3)			X	
attending a legal service plan (2)				X
attitudes must be met				X
by learning knowledge				X
complex works				X
contact the topic				X
cost which was saved (2)				X
cultivating desalinization plants				X
customer psychology (3)				X
efforts which have been done	X			
extensively curtail				X
fit the desires of them			X	X
fit the esteem and self-actualization needs (3)				X
gather water				X
graduation commitment (3)		X		
how to result the problem		X		X
huge abilities				X
in a few years ago				X
keep the whole assets (2)				X
make the good influence on the world (3)	X		X	
makes no difference between				X
manager the property		X	X	X
memberships of the plans		X	X	
middle level persons				X
open the computer	X			X

TABLE (Continued)

Collocation Cross-Classification for Each Student -- PO

Collocation	Lex Transfer	Phon	Gram	Sem
original beginning of spenthrift trust				X
price strategy (3)		X	X	
put an alternate for trustee (2)				X
save the income tax (3)				X
saving taxation				X
successful once in the past				X
takes it for example				X
the act in law (2)			X	
the data of consumer behavior (4)			X	
the different education style				X
the government appealed people			X	
the market can be diversified into				X
the market was diversified in				X
the method of saving tax (2)				X
the reflection from competitor		X		X
the routine law service (2)				X
the salary notice				X
to adapt			X	
to save taxation				X
to survival the growing population			X	X
to use their political rights				X
worth to pay			X	

TABLE

Collocation Cross-Classification for Each Student -- CH

Collocation	Lex Transfer	Phon	Gram	Sem
abundant working experience (3)				X
accelerating old	X		X	
accept any interrupts				X
arrange workers during the rush hour	X			X
assistant its growth		X	X	X
because of the gap of age			X	
better health technique (2)				X
both physical and mind condition				X
catch up the step of the changing world (2)				X
change to diversified				X
companies should do some change	X			
conscious of moral		X	X	
declining the rate of sick				X
degree of how			X	X
distinct themselves from others		X		X
distinctive grope of people		X		

TABLE (Continued)

Collocation Cross-Classification for Each Student -- CH

Collocation	Lex Transfer	Phon	Gram	Sem
extending life expectation	X			X
familiar the new skill			X	
feel tied about their finance			X	
full of information's era			X	
gratify what they think				X
harmony working environment		X	X	
have very poor shapes			X	X
increasing number of older generation				X
jobbing retaining			X	
less responsibilities				X
losing your friend's time				X
mainstreams of the economy		X		X
make them more worth than younger			X	
market share rate				X
moral and freedom		X	X	
moral conscious		X		X
most potential environmental disaster	X			
older adults are less absent			X	
our middle life				X
people working after a certain years				X
physical failing		X		X
retirees' health is accelerating failure				X
revive the irreversible damages		X		X
satisfy their value				X
search jobs			X	
supply has been over demand				X
take a lot of advantages by				X
target marketing is wildly used		X		
the aging era				X
the average life of retirees				X
the change of the demographic			X	
the expedition of the human life				X
the left 35 years				X
the main strain of the society		X		X
the rate of the consumers				X
the supply have been over demand				X
their physical fail		X		X
their too cautious			X	X
they still own numerous knowledge				X
to adapt			X	
to adopt ourselves		X		
to close nature			X	
to close the nature			X	
to demise the conflicts				X
to grasp knowledge				X
to opt the concentrated marketing			X	

TABLE (Continued)

Collocation Cross-Classification for Each Student -- CH

Collocation	Lex Transfer	Phon	Gram	Sem
to replace their jobs				X
who are above 55				X
will improve with getting old	X			X
worthy to do it			X	
years of their rest life				X

TABLE

Collocation Cross-Classification for Each Student -- LU

Collocation	Lex Transfer	Phon	Gram	Sem
a long time trip	X			
a number of evidence				X
adopted past experiences		X		
advertisements' budget			X	
analyze method			X	
as can be seen			X	
be close the nature			X	
budget of advertising			X	
business portion				X
complain their performances			X	
completely realize				X
crowded with the contradictions				X
food, clothes, and shelters			X	
good performance of sale			X	
housing, clothing, and eating				X
in the rest year				X
keep in a small area	X			X
largely decreased				X
lives' quality			X	
living space				X
making us keep in a small area	X			X
market characteristic				X
more market share				X
pay more salary				X
pay practice				X
receive new data				X
satisfy what the employers need			X	X
satisfy what they need			X	X
the degree of how			X	X
the degree to which how			X	X
the female aspect	X			X
the interior of company structure		X		X

TABLE (Continued)

Collocation Cross-Classification for Each Student -- LU

Collocation	Lex Transfer	Phon	Gram	Sem
the level of how			X	X
the most potential market	X		X	X
the portion of female user				X
the products are various		X	X	
the survey should be remade	X			X
to close the nature			X	
to distinct those who use this product		X		X
to fit any kind of circumstances				X
to fit the need				X
to gain more competitive				X
to prove my guess				X
to realize how the market is				X
unbenefit behavior				X
war of marketing			X	

TABLE

Collocation Cross-Classification for Each Student -- DI

Collocation	Lex Transfer	Phon	Gram	Sem
a large part of the people				X
a scientist raised a technique				X
a survey made by	X			
a survey was made	X			
adjust themselves		X		X
an intense impression				X
attention is paid on them			X	
decrease the rate of crime				X
definite customers			X	
fewer-children marriages	X			X
huge different action				X
luxurious items		X	X	
matches consumers needs				X
nowadays large numbers				X
purchasing abilities				X
raise its sales				X
save the situation from worsening	X			X
the aging generation				X
the businessmen market		X	X	
the data of statistic figures			X	
the product figure				X
the role they act				X
the survey made by	X			
to adapt			X	
to cater for popular tastes			X	

TABLE Continued)

Collocation Cross-Classification for Each Student -- DI

Collocation	Lex Transfer	Phon	Gram	Sem
to go toward this direction			X	
to lack well support				X
value concepts				X
win the success				X

TABLE

Collocation Cross-Classification for Each Student -- GO

Collocation	Lex Transfer	Phon	Gram	Sem
a certain time range	X			X
a number of money				X
accomplish difficultly			X	
adds its advertising budget			X	
all styles of works			X	
As could be seen			X	
assign time recording				X
attendance cards				X
avoid being out of the trend				X
business class				X
care price			X	
care the price			X	
contain the total number of hours				X
cost losses				X
costs can contain				X
ethic in value concepts				X
experiments had previously been made	X			
focus on the business class				X
from business class to student class				X
fully important				X
goods with low prices			X	
heighten his quality of life				X
hot working time				X
In past time				X
it is to noted that			X	
losses of time cost				X
make a conclusion	X			X
make a market survey	X			
make a meeting	X			X
market share falls down			X	X
market share has falled down			X	X
market share has fallen down			X	X
money society				X
open the monitor and computer	X			X

TABLE (Continued)

Collocation Cross-Classification for Each Student -- GO

Collocation	Lex Transfer	Phon	Gram	Sem
quit this prescription				X
rise moral standards		X		X
scan the two systems carefully				X
so huge abilities				X
stability is reducing				X
suffered from a significant improvement				X
the chances of learning knowledge				X
their working time				X
to clarify these doubts				X
to pay more salaries				X
To sum			X	
wastes of sources		X		X
work in hot time				X

TABLE

Collocation Cross-Classification for Each Student -- LY

Collocation	Lex Transfer	Phon	Gram	Sem
almost two-third		X		
apply this theory into			X	
are tend to in favor of	X			
build up their own features				X
compile them into different categories				X
consuming behavior		X	X	
consuming behaviors		X	X	
consuming type		X	X	
decline in relevant to		X		X
developing space				X
differentiate percentage		X		X
difficult portion for marketing				X
do not in favor of	X			
dollars are relevant weak		X		X
dollars are relevantly weak		X		X
dollars keep falling down			X	X
dollars relevant weak		X		X
easy to aware			X	
economical progress		X	X	
economical prosperity		X	X	
economical stability		X	X	
economical trend		X	X	
faithful on the same products				X
fill of		X		
filled with explosive technology				X

TABLE (Continued)

Collocation Cross-Classification for Each Student -- LY

Collocation	Lex Transfer	Phon	Gram	Sem
forever winner or loser	X			X
grand amount of currency exchange				X
high class cloth				X
intensive competition		X		X
key of achieving market success			X	
key of success			X	
key of the good strategies			X	
large amount of dollars				X
London's share is falling down			X	X
long options for hedging				X
lower down the price			X	X
more dollars relative than			X	
most potential buyer	X			X
On the opposite				X
over four-five	X	X		X
pass through in front of us			X	X
pose a significant role for				X
products are pointed				X
push the US dollar's rising			X	X
raise the crimes committing rate				X
relative with people's life		X		X
replans its company goal				X
rescue the food systems				X
rise the interest rate		X		X
rival society				X
safe heaven in a sea of global trouble		X		X
scientists make a research	X			
severe competition	X			X
since six months ago				X
spend so many time				X
target its new market into students			X	X
the backup of the greenback				X
the dollars will not always keep down	X			X
the most possible buyers	X			X
the most potential buyers	X			X
the rescue for the morality crisis				X
three-four of consumers	X	X		
three-four of students	X	X		
to accommodate the disaster				X
to adapt			X	
to adjust itself		X		X
to avoid loss money			X	
translating the dollars				X
twice times	X			X
US dollars might keep weak	X			X
we made an investigation	X			X

TABLE (Continued)

Collocation Cross-Classification for Each Student -- LY

Collocation	Lex Transfer	Phon	Gram	Sem
women occupy three-four of consumers				X

TABLE

Collocation Cross-Classification for Each Student -- WI

Collocation	Lex Transfer	Phon	Gram	Sem
a great number of money				X
a little slipping down			X	X
accepted broadly				X
add...once at a time	X	X		X
adjusting themselves to survive		X		X
As can been seen			X	
consumer of women			X	
consuming ability				X
dramatically going down			X	X
increase the distance				X
extremely possible				X
falses at the beginning				X
figure out its potentiality		X	X	
further realize				X
greatly possible				X
in the unknown future				X
installed over times		X		X
market segments decreased				X
one command may be made	X			X
points that			X	
release the poverty problem		X		X
rooted concepts				X
solved thoroughly				X
special consumer potentiality		X	X	
still slipping down			X	X
the more possible				X
three-fourth students'		X	X	
twice times more	X			X
value concepts				X
widen his knowledge				X

APPENDIX E

CONTINGENCY TABLES OF COLLOCATION TOKENS
FOR EACH STUDENT

Contingency Table of Collocation Tokens for CA: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	4	54	58
Semantic -- N	8	15	23
Column Total	12	69	81

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	8.60	49.41	58
Semantic -- N	3.40	19.59	23
Column Total	12.00	69.00	81

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	3.02	0.34	3.36
Semantic -- N	4.92	1.32	6.24
Column Total	7.94	1.66	9.60

Contingency Table of Collocation Tokens for CA: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2	19	21
Phonology -- N	10	50	60
Column Total	12	69	81

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	3.11	17.89	21
Phonology -- N	8.89	51.11	60
Column Total	12.00	69.00	81

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.83	0.02	0.85
Phonology -- N	0.04	0.05	0.09
Column Total	0.87	0.07	0.94

Contingency Table of Collocation Tokens for CA: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	14	7	21
Phonology -- N	44	16	60
Column Total	58	23	81

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	15.04	5.96	21
Phonology -- N	42.96	17.04	60
Column Total	58.00	23.00	81

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.16	0.05	0.21
Phonology -- N	0.00	0.14	0.14
Column Total	0.16	0.19	0.35

Contingency Table of Collocation Tokens for CA: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	12	12
Grammar -- N	12	57	69
Column Total	12	69	81

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.78	10.22	12
Grammar -- N	10.22	58.78	69
Column Total	12.00	69.00	81

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.92	0.16	3.08
Grammar -- N	0.16	0.09	0.25
Column Total	3.08	0.25	3.33

Contingency Table of Collocation Tokens for CA: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	2	10	12
Grammar -- N	56	13	69
Column Total	58	23	81

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	8.59	3.41	12
Grammar -- N	49.41	19.59	69
Column Total	58.00	23.00	81

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	5.85	10.89	16.74
Grammar -- N	0.75	2.57	3.32
Column Total	6.60	13.46	20.06

Contingency Table of Collocation Tokens for CA: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	2	10	12
Grammar -- N	19	50	69
Column Total	21	60	81

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	3.11	8.89	12
Grammar -- N	17.89	51.11	69
Column Total	21.00	60.00	81

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.83	0.04	0.87
Grammar -- N	0.02	0.05	0.07
Column Total	0.85	0.09	0.94

Contingency Table of Collocation Tokens for CH: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	5	56	61
Semantic -- N	4	27	31
Column Total	9	83	92

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	5.97	55.03	61
Semantic -- N	3.03	27.97	31
Column Total	9.00	83.00	92

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.36	0.00	0.36
Semantic -- N	0.07	0.08	0.15
Column Total	0.43	0.08	0.51

Contingency Table of Collocation Tokens for CH: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	22	22
Phonology -- N	9	61	70
Column Total	9	83	92

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.15	19.85	22
Phonology -- N	6.85	63.15	70
Column Total	9.00	83.00	92

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	3.27	0.14	3.41
Phonology -- N	0.40	0.11	0.51
Column Total	3.67	0.25	3.92

Contingency Table of Collocation Tokens for CH: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	12	10	22
Phonology -- N	49	21	70
Column Total	61	31	92

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	14.59	7.41	22
Phonology -- N	46.41	23.59	70
Column Total	61.00	31.00	92

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.65	0.59	1.24
Phonology -- N	0.10	0.40	0.50
Column Total	0.75	0.99	1.74

Contingency Table of Collocation Tokens for CH: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1	28	29
Grammar -- N	8	55	63
Column Total	9	83	92

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.84	26.16	29
Grammar -- N	6.16	56.84	63
Column Total	9.00	83.00	92

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.92	0.07	1.99
Grammar -- N	0.29	0.09	0.38
Column Total	2.21	0.16	2.37

Contingency Table of Collocation Tokens for CH: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	4	25	29
Grammar -- N	57	6	63
Column Total	61	31	92

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	19.23	9.77	29
Grammar -- N	41.77	21.23	63
Column Total	61.00	31.00	92

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	12.87	22.20	35.07
Grammar -- N	5.19	11.65	16.84
Column Total	18.06	33.85	51.91

Contingency Table of Collocation Tokens for CH: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	8	21	29
Grammar -- N	14	49	63
Column Total	22	70	92

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	6.93	22.07	29
Grammar -- N	15.07	47.93	63
Column Total	22.00	70.00	92

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.05	0.11	0.16
Grammar -- N	0.16	0.01	0.17
Column Total	0.21	0.12	0.33

Contingency Table of Collocation Tokens for DI: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	3	29	32
Semantic -- N	7	11	18
Column Total	10	40	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	6.4	25.6	32
Semantic -- N	3.6	14.4	18
Column Total	10.0	40.0	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	2.38	0.33	2.71
Semantic -- N	2.33	1.06	3.39
Column Total	4.71	1.39	6.10

Contingency Table of Collocation Tokens for DI: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	4	4
Phonology -- N	10	36	46
Column Total	10	40	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.8	3.2	4
Phonology -- N	9.2	36.8	46
Column Total	10.0	40.0	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.11	0.03	2.14
Phonology -- N	0.01	0.04	0.05
Column Total	2.12	0.07	2.19

Contingency Table of Collocation Tokens for DI: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1	3	4
Phonology -- N	31	15	46
Column Total	32	18	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	2.56	1.44	4
Phonology -- N	29.44	16.56	46
Column Total	32.00	18.00	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.66	0.78	2.44
Phonology -- N	0.04	0.25	0.29
Column Total	1.70	1.03	2.73

Contingency Table of Collocation Tokens for DI: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	11	11
Grammar -- N	10	29	39
Column Total	10	40	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.2	8.8	11
Grammar -- N	7.8	31.2	39
Column Total	10.0	40.0	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	3.31	0.33	3.64
Grammar -- N	0.37	0.23	0.60
Column Total	3.68	0.56	4.24

Contingency Table of Collocation Tokens for DI: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	0	11	11
Grammar -- N	32	7	39
Column Total	32	18	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	7.04	3.96	11
Grammar -- N	24.96	14.04	39
Column Total	32.00	18.00	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	8.08	10.80	18.88
Grammar -- N	1.71	4.05	5.76
Column Total	9.79	14.85	24.64

Contingency Table of Collocation Tokens for DI: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	3	8	11
Grammar -- N	1	38	39
Column Total	4	46	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.88	10.12	11
Grammar -- N	3.12	35.88	39
Column Total	4.00	46.00	50.00

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	2.98	0.68	3.66
Grammar -- N	2.20	0.07	2.27
Column Total	5.18	0.75	5.93

Contingency Table of Collocation Tokens for GO: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	5	47	52
Semantic -- N	3	16	19
Column Total	8	63	71

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	5.86	46.14	52
Semantic -- N	2.14	16.86	19
Column Total	8.00	63.00	71

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.32	0.00	0.32
Semantic -- N	0.06	0.11	0.17
Column Total	0.38	0.11	0.49

Contingency Table of Collocation Tokens for GO: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	2	2
Phonology -- N	8	61	69
Column Total	8	63	71

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.23	1.77	2
Phonology -- N	7.77	61.23	69
Column Total	8.00	63.00	71

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.33	0.04	2.38
Phonology -- N	0.01	0.01	0.02
Column Total	2.34	0.05	2.40

Contingency Table of Collocation Tokens for GO: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	2	0	2
Phonology -- N	50	19	69
Column Total	52	19	71

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.46	0.54	2
Phonology -- N	50.54	18.46	69
Column Total	52.00	19.00	71

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.00	2.00	2.00
Phonology -- N	0.02	0.00	0.02
Column Total	0.02	2.00	2.02

Contingency Table of Collocation Tokens for GO: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	20	20
Grammar -- N	8	43	51
Column Total	8	63	71

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.25	17.75	20
Grammar -- N	5.75	45.25	51
Column Total	8.00	63.00	71

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	3.37	0.17	3.54
Grammar -- N	0.53	0.17	0.70
Column Total	3.90	0.34	4.24

Contingency Table of Collocation Tokens for GO: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	4	16	20
Grammar -- N	48	3	51
Column Total	52	19	71

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	14.65	5.35	20
Grammar -- N	37.35	13.65	51
Column Total	52.00	19.00	71

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	8.48	19.24	27.72
Grammar -- N	2.76	9.10	11.86
Column Total	11.24	28.34	39.58

Contingency Table of Collocation Tokens for GO: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0	20	20
Grammar -- N	2	49	51
Column Total	2	69	71

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.56	19.44	20
Grammar -- N	1.44	49.56	51
Column Total	2.00	69.00	71

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	2.01	0.00	2.01
Grammar -- N	0.00	0.02	0.02
Column Total	2.01	0.02	2.03

Contingency Table of Collocation Tokens for LU: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	8	40	48
Semantic -- N	1	23	24
Column Total	9	63	72

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	6	42	48
Semantic -- N	3	21	24
Column Total	9	63	72

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.38	0.15	0.53
Semantic -- N	2.08	0.11	2.19
Column Total	2.46	0.26	2.71

Contingency Table of Collocation Tokens for LU: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	7	7
Phonology -- N	9	56	65
Column Total	9	63	72

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.87	6.13	7
Phonology -- N	8.13	56.87	65
Column Total	9.00	63.00	72

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.16	0.02	2.18
Phonology -- N	0.02	0.03	0.05
Column Total	2.18	0.05	2.23

Contingency Table of Collocation Tokens for LU: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	3	4	7
Phonology -- N	45	20	65
Column Total	48	24	72

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	4.67	2.33	7
Phonology -- N	43.33	21.67	65
Column Total	48.00	24.00	72

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.01	0.58	1.59
Phonology -- N	0.03	0.22	0.25
Column Total	1.04	0.80	1.84

Contingency Table of Collocation Tokens for LU: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	3	30	33
Grammar -- N	6	33	39
Column Total	9	63	72

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	4.13	28.87	33
Grammar -- N	4.87	34.13	39
Column Total	9.00	63.00	72

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0.64	0.01	0.65
Grammar -- N	0.08	0.08	0.16
Column Total	0.72	0.09	0.81

Contingency Table of Collocation Tokens for LU: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	12	21	33
Grammar -- N	36	3	39
Column Total	48	24	72

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	22	11	33
Grammar -- N	26	13	39
Column Total	48	24	72

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	5.01	8.20	13.21
Grammar -- N	3.47	8.48	11.95
Column Total	8.48	16.68	25.16

Contingency Table of Collocation Tokens for LU: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	2	31	33
Grammar -- N	5	34	39
Column Total	7	65	72

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	3.21	29.79	33
Grammar -- N	3.79	35.21	39
Column Total	7.00	65.00	72

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.91	0.02	0.93
Grammar -- N	0.13	0.08	0.21
Column Total	1.04	0.10	1.14

Contingency Table of Collocation Tokens for LY: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	12	45	57
Semantic -- N	6	22	28
Column Total	18	67	85

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	12.07	44.93	57
Semantic -- N	5.93	22.07	28
Column Total	18.00	67.00	85

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.03	0.01	0.04
Semantic -- N	0.03	0.01	0.04
Column Total	0.06	0.02	0.08

Contingency Table of Collocation Tokens for LY: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	3	24	27
Phonology -- N	15	43	58
Column Total	18	67	85

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	5.72	21.28	27
Phonology -- N	12.28	45.72	58
Column Total	18.00	67.00	85

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	1.81	0.23	2.04
Phonology -- N	0.40	0.23	0.63
Column Total	2.21	0.46	2.67

Contingency Table of Collocation Tokens for LY: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	13	14	27
Phonology -- N	44	14	58
Column Total	57	28	85

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	18.11	8.89	27
Phonology -- N	38.89	19.11	58
Column Total	57.00	28.00	85

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.73	2.39	4.12
Phonology -- N	0.55	1.64	2.19
Column Total	2.28	4.03	6.31

Contingency Table of Collocation Tokens for LY: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1	26	27
Grammar -- N	17	41	58
Column Total	18	67	85

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	5.72	21.28	27
Grammar -- N	12.28	45.72	58
Column Total	18.00	67.00	85

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	4.76	0.84	5.60
Grammar -- N	1.45	0.59	2.04
Column Total	6.21	1.43	7.64

Contingency Table of Collocation Tokens for LY: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	8	19	27
Grammar -- N	49	9	58
Column Total	57	28	85

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	18.11	8.89	27
Grammar -- N	38.89	19.11	58
Column Total	57.00	28.00	85

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	6.21	10.37	16.58
Grammar -- N	2.37	5.89	8.26
Column Total	8.58	16.26	24.84

Contingency Table of Collocation Tokens for LY: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	9	18	27
Grammar -- N	18	40	58
Column Total	27	58	85

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	8.58	18.42	27
Grammar -- N	18.42	39.58	58
Column Total	27.00	58.00	85

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.00	0.05	0.05
Grammar -- N	0.05	0.00	0.05
Column Total	0.05	0.05	0.10

Contingency Table of Collocation Tokens for PO: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	4	50	54
Semantic -- N	1	20	21
Column Total	5	70	75

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	3.6	50.4	54
Semantic -- N	1.4	19.6	21
Column Total	5.0	70.0	75

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.00	0.02	0.02
Semantic -- N	0.58	0.00	0.58
Column Total	0.58	0.02	0.60

Contingency Table of Collocation Tokens for PO: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	12	12
Phonology -- N	5	58	63
Column Total	5	70	75

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.8	11.2	12
Phonology -- N	4.2	58.8	63
Column Total	5.0	70.0	75

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.11	0.01	2.12
Phonology -- N	0.02	0.03	0.05
Column Total	2.13	0.04	2.17

Contingency Table of Collocation Tokens for PO: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	5	7	12
Phonology -- N	49	14	63
Column Total	54	21	75

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	8.64	3.36	12
Phonology -- N	45.36	17.64	63
Column Total	54.00	21.00	75

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.98	2.93	4.91
Phonology -- N	0.22	0.97	1.19
Column Total	2.20	3.90	6.10

Contingency Table of Collocation Tokens for PO: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	21	21
Grammar -- N	5	49	54
Column Total	5	70	75

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.4	19.6	21
Grammar -- N	3.6	50.4	54
Column Total	5.0	70.0	75

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.58	0.04	2.62
Grammar -- N	0.22	0.07	0.29
Column Total	2.80	0.11	2.91

Contingency Table of Collocation Tokens for PO: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	4	17	21
Grammar -- N	50	4	54
Column Total	54	21	75

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	15.12	5.88	21
Grammar -- N	38.88	15.12	54
Column Total	54.00	21.00	75

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	8.93	19.18	28.11
Grammar -- N	2.90	8.93	11.83
Column Total	11.83	28.11	39.94

Contingency Table of Collocation Tokens for PO: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	5	16	21
Grammar -- N	7	47	54
Column Total	12	63	75

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	3.36	17.64	21
Grammar -- N	8.64	45.36	54
Column Total	12.00	63.00	75

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.39	0.26	0.65
Grammar -- N	0.53	0.03	0.56
Column Total	0.92	0.29	1.21

Contingency Table of Collocation Tokens for WI: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	3	29	32
Semantic -- N	0	13	13
Column Total	3	42	45

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	2.13	29.87	32
Semantic -- N	0.87	12.13	13
Column Total	3.00	42.00	45

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.06	0.06	0.12
Semantic -- N	2.16	0.01	2.17
Column Total	2.22	0.07	2.29

Contingency Table of Collocation Tokens for WI: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	1	10	11
Phonology -- N	2	32	34
Column Total	3	42	45

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.73	10.27	11
Phonology -- N	2.27	31.73	34
Column Total	3.00	42.00	45

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.07	0.06	0.13
Phonology -- N	0.26	0.00	0.26
Column Total	0.33	0.06	0.39

Contingency Table of Collocation Tokens for WI: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	4	7	11
Phonology -- N	28	6	34
Column Total	32	13	45

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	7.82	3.18	11
Phonology -- N	24.18	9.82	34
Column Total	32.00	13.00	45

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	2.39	3.47	5.86
Phonology -- N	0.45	1.90	2.35
Column Total	2.84	5.37	8.21

Contingency Table of Collocation Tokens for WI: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	20	20
Grammar -- N	3	22	25
Column Total	3	42	45

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.33	18.67	20
Grammar -- N	1.67	23.33	25
Column Total	3.00	42.00	45

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.52	0.04	2.56
Grammar -- N	0.42	0.14	0.56
Column Total	2.94	0.18	3.11

Contingency Table of Collocation Tokens for WI: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	7	13	20
Grammar -- N	25	0	25
Column Total	32	13	45

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	14.22	5.78	20
Grammar -- N	17.78	7.22	25
Column Total	32.00	13.00	45

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	4.19	7.82	12.01
Grammar -- N	2.54	8.26	10.80
Column Total	6.73	16.08	22.81

Contingency Table of Collocation Tokens for WI: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	7	13	20
Grammar -- N	4	21	25
Column Total	11	34	45

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	4.89	15.11	20
Grammar -- N	6.11	18.89	25
Column Total	11.00	34.00	45

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.53	0.45	0.98
Grammar -- N	1.12	0.14	1.26
Column Total	1.65	0.59	2.24

APPENDIX F

CONTINGENCY TABLES OF COLLOCATION TYPES
FOR EACH STUDENT

Contingency Table of Collocation Types for CA: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	3	39	42
Semantic -- N	6	13	19
Column Total	9	52	61

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	6.20	35.80	42
Semantic -- N	2.80	16.20	19
Column Total	9.00	52.00	61

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	2.21	0.20	2.41
Semantic -- N	2.59	0.85	3.44
Column Total	4.80	1.05	5.85

Contingency Table of Collocation Types for CA: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	1	13	14
Phonology -- N	8	39	47
Column Total	9	52	61

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.07	11.93	14
Phonology -- N	6.93	40.07	47
Column Total	9.00	52.00	61

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	1.19	0.03	1.21
Phonology -- N	0.04	0.06	0.11
Column Total	1.23	0.09	1.32

Contingency Table of Collocation Types for CA: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	9	5	14
Phonology -- N	33	14	47
Column Total	42	19	61

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	9.64	4.36	14
Phonology -- N	32.36	14.64	47
Column Total	42.00	19.00	61

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.13	0.01	0.13
Phonology -- N	0.01	0.08	0.10
Column Total	0.14	0.09	0.23

Contingency Table of Collocation Types for CA: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	9	9
Grammar -- N	9	43	52
Column Total	9	52	61

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.33	7.67	9
Grammar -- N	7.67	44.33	52
Column Total	9.00	52.00	61

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.52	0.09	2.60
Grammar -- N	0.09	0.07	0.17
Column Total	2.61	0.16	2.77

Contingency Table of Collocation Types for CA: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	2	7	9
Grammar -- N	40	12	52
Column Total	42	19	61

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	6.20	2.80	9
Grammar -- N	35.80	16.20	52
Column Total	42.00	19.00	61

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	3.56	4.87	8.43
Grammar -- N	0.38	1.36	1.74
Column Total	3.94	6.23	10.17

Contingency Table of Collocation Types for CA: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	1	8	9
Grammar -- N	13	39	52
Column Total	14	47	61

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	2.07	6.93	9
Grammar -- N	11.93	40.07	52
Column Total	14.00	47.00	61

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	1.19	0.05	1.23
Grammar -- N	0.02	0.06	0.09
Column Total	1.21	0.11	1.320712

Contingency Table of Collocation Types for CH: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	4	41	45
Semantic -- N	4	24	28
Column Total	8	65	73

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	4.93	40.07	45
Semantic -- N	3.07	24.93	28
Column Total	8.00	65.00	73

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.42	0.00	0.42
Semantic -- N	0.06	0.08	0.14
Column Total	0.48	0.08	0.56

Contingency Table of Collocation Types for CH: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	15	15
Phonology -- N	8	50	58
Column Total	8	65	73

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	1.64	13.36	15
Phonology -- N	6.36	51.64	58
Column Total	8.00	65.00	73

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.79	0.10	2.89
Phonology -- N	0.21	0.08	0.29
Column Total	3.00	0.18	3.18

Contingency Table of Collocation Types for CH: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	8	7	15
Phonology -- N	37	21	58
Column Total	45	28	73

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	9.25	5.75	15
Phonology -- N	35.75	22.25	58
Column Total	45.00	28.00	73

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.33	0.09	0.42
Phonology -- N	0.02	0.14	0.16
Column Total	0.35	0.23	0.58

Contingency Table of Collocation Types for CH: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1	24	25
Grammar -- N	7	41	48
Column Total	8	65	73

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.74	22.26	25
Grammar -- N	5.26	42.74	48
Column Total	8.00	65.00	73

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.83	0.07	1.90
Grammar -- N	0.30	0.11	0.41
Column Total	2.13	0.18	2.31

Contingency Table of Collocation Types for CH: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	3	22	25
Grammar -- N	42	6	48
Column Total	45	28	73

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	15.41	9.59	25
Grammar -- N	29.59	18.41	48
Column Total	45.00	28.00	73

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	10.82	14.79	25.61
Grammar -- N	4.79	9.06	13.84
Column Total	15.61	23.85	39.46

Contingency Table of Collocation Types for CH: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	5	20	25
Grammar -- N	10	38	48
Column Total	15	58	73

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	5.14	19.86	25
Grammar -- N	9.86	38.14	48
Column Total	15.00	58.00	73

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.08	0.01	0.09
Grammar -- N	0.01	0.01	0.02
Column Total	0.09	0.02	0.11

Contingency Table of Collocation Types for DI: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	2	18	20
Semantic -- N	3	8	11
Column Total	5	26	31

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	3.23	16.77	20
Semantic -- N	1.77	9.23	11
Column Total	5.00	26.00	31

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.92	0.03	0.95
Semantic -- N	0.30	0.32	0.62
Column Total	1.22	0.35	1.57

Contingency Table of Collocation Types for DI: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	3	3
Phonology -- N	5	23	28
Column Total	5	26	31

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.48	2.52	3
Phonology -- N	4.52	23.48	28
Column Total	5.00	26.00	31

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.00	0.00	2.00
Phonology -- N	0.00	0.04	0.04
Column Total	2.00	0.04	2.04

Contingency Table of Collocation Types for DI: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1	2	3
Phonology -- N	19	9	28
Column Total	20	11	31

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.94	1.06	3
Phonology -- N	18.06	9.94	28
Column Total	20.00	11.00	31

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.06	0.18	1.24
Phonology -- N	0.01	0.21	0.22
Column Total	1.07	0.39	1.46

Contingency Table of Collocation Types for DI: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	8	8
Grammar -- N	5	18	23
Column Total	5	26	31

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.29	6.71	8
Grammar -- N	3.71	19.29	23
Column Total	5.00	26.00	31

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.48	0.10	2.58
Grammar -- N	0.17	0.16	0.33
Column Total	2.65	0.26	2.91

Contingency Table of Collocation Types for DI: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	0	8	8
Grammar -- N	20	3	23
Column Total	20	11	31

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	5.16	2.84	8
Grammar -- N	14.84	8.16	23
Column Total	20.00	11.00	31

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	6.21	7.65	13.86
Grammar -- N	1.46	3.93	5.39
Column Total	7.67	11.58	19.25

Contingency Table of Collocation Types for DI: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	2	6	8
Grammar -- N	1	22	23
Column Total	3	28	31

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.77	7.23	8
Grammar -- N	2.23	20.77	23
Column Total	3.00	28.00	31

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.69	0.41	1.10
Grammar -- N	1.33	0.03	1.36
Column Total	2.02	0.44	2.46

Contingency Table of Collocation Types for GO: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	4	32	36
Semantic -- N	2	10	12
Column Total	6	42	48

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	4.5	31.5	36
Semantic -- N	1.5	10.5	12
Column Total	6.0	42.0	48

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.22	0.00	0.22
Semantic -- N	0.00	0.10	0.10
Column Total	0.22	0.10	0.32

Contingency Table of Collocation Types for GO: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	2	2
Phonology -- N	6	40	46
Column Total	6	42	48

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.25	1.75	2
Phonology -- N	5.75	40.25	46
Column Total	6.00	42.00	48

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.25	0.04	2.29
Phonology -- N	0.01	0.01	0.02
Column Total	2.26	0.05	2.31

Contingency Table of Collocation Types for GO: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	2	0	2
Phonology -- N	34	12	46
Column Total	36	12	48

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.5	0.5	2
Phonology -- N	34.5	11.5	46
Column Total	36.0	12.0	48

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.00	2.00	2.00
Phonology -- N	0.03	0.00	0.03
Column Total	0.03	2.00	2.03

Contingency Table of Collocation Types for GO: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	13	13
Grammar -- N	6	29	35
Column Total	6	42	48

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.63	11.38	13
Grammar -- N	4.37	30.62	35
Column Total	6.00	42.00	48

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.78	0.11	2.89
Grammar -- N	0.29	0.15	0.44
Column Total	3.07	0.26	3.33

Contingency Table of Collocation Types for GO: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	3	10	13
Grammar -- N	33	2	35
Column Total	36	12	48

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	9.75	3.25	13
Grammar -- N	26.25	8.75	35
Column Total	36.00	12.00	48

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	5.39	12.02	17.41
Grammar -- N	1.49	6.01	7.50
Column Total	6.88	18.03	24.91

Contingency Table of Collocation Types for GO: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0	13	13
Grammar -- N	2	33	35
Column Total	2	46	48

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.54	12.46	13
Grammar -- N	1.46	33.54	35
Column Total	2.00	46.00	48

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	2.00	0.00	2.00
Grammar -- N	0.00	0.03	0.03
Column Total	2.00	0.03	2.03

Contingency Table of Collocation Types for LU: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	5	28	33
Semantic -- N	1	13	14
Column Total	6	41	47

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	4.21	28.79	33
Semantic -- N	1.79	12.21	14
Column Total	6.00	41.00	47

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.02	0.06	0.08
Semantic -- N	0.93	0.00	0.93
Column Total	0.95	0.06	1.01

Contingency Table of Collocation Types for LU: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	4	4
Phonology -- N	6	37	43
Column Total	6	41	47

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.51	3.49	4
Phonology -- N	5.49	37.51	43
Column Total	6.00	41.00	47

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.00	0.00	2.00
Phonology -- N	0.00	0.03	0.03
Column Total	2.00	0.03	2.03

Contingency Table of Collocation Types for LU: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	2	2	4
Phonology -- N	31	12	43
Column Total	33	14	47

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	2.81	1.19	4
Phonology -- N	30.19	12.81	43
Column Total	33.00	14.00	47

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.61	0.08	0.69
Phonology -- N	0.00	0.13	0.13
Column Total	0.61	0.21	0.82

Contingency Table of Collocation Types for LU: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1	17	18
Grammar -- N	5	24	29
Column Total	6	41	47

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.30	15.70	18
Grammar -- N	3.70	25.30	29
Column Total	6.00	41.00	47

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1.41	0.04	1.45
Grammar -- N	0.17	0.13	0.30
Column Total	1.58	0.17	1.75

Contingency Table of Collocation Types for LU: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	6	12	18
Grammar -- N	27	1	28
Column Total	33	13	46

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	12.91	5.09	18
Grammar -- N	20.09	7.91	28
Column Total	33.00	13.00	46

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	4.25	8.09	12.34
Grammar -- N	2.05	6.94	8.99
Column Total	6.30	15.03	21.33

Contingency Table of Collocation Types for LU: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	1	17	18
Grammar -- N	3	26	29
Column Total	4	43	47

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	1.53	16.47	18
Grammar -- N	2.47	26.53	29
Column Total	4.00	43.00	47

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.70	0.00	0.70
Grammar -- N	0.00	0.04	0.04
Column Total	0.70	0.04	0.74

Contingency Table of Collocation Types for LY: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	12	41	53
Semantic -- N	5	21	26
Column Total	17	62	79

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	11.41	41.60	53
Semantic -- N	5.59	20.40	26
Column Total	17.00	62.00	79

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.00	0.03	0.03
Semantic -- N	0.21	0.00	0.21
Column Total	0.21	0.03	0.24

Contingency Table of Collocation Types for LY: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	3	22	25
Phonology -- N	14	40	54
Column Total	17	62	79

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	5.38	19.62	25
Phonology -- N	11.62	42.38	54
Column Total	17.00	62.00	79

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	1.54	0.18	1.72
Phonology -- N	0.30	0.20	0.50
Column Total	1.84	0.38	2.22

Contingency Table of Collocation Types for LY: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	12	13	25
Phonology -- N	41	13	54
Column Total	53	26	79

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	16.77	8.23	25
Phonology -- N	36.23	17.77	54
Column Total	53.00	26.00	79

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	1.66	2.22	3.87
Phonology -- N	0.50	1.56	2.07
Column Total	2.16	3.78	5.94

Contingency Table of Collocation Types for LY: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	1	24	25
Grammar -- N	16	38	54
Column Total	17	62	79

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	5.38	19.62	25
Grammar -- N	11.62	42.38	54
Column Total	17.00	62.00	79

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	4.43	0.77	5.19
Grammar -- N	1.29	0.56	1.86
Column Total	5.72	1.33	7.05

Contingency Table of Collocation Types for LY: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	7	18	25
Grammar -- N	46	8	54
Column Total	53	26	79

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	16.77	8.23	25
Grammar -- N	36.23	17.77	54
Column Total	53.00	26.00	79

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	6.29	10.45	16.74
Grammar -- N	2.37	5.94	8.31
Column Total	8.66	16.39	25.05

Contingency Table of Collocation Types for LY: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	8	17	25
Grammar -- N	17	37	54
Column Total	25	54	79

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	7.91	17.09	25
Grammar -- N	17.09	36.91	54
Column Total	25.00	54.00	79

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.02	0.02	0.04
Grammar -- N	0.02	0.00	0.02
Column Total	0.04	0.02	0.06

Contingency Table of Collocation Types for PO: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	2	37	39
Semantic -- N	1	10	11
Column Total	3	47	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	2.34	36.66	39
Semantic -- N	0.66	10.34	11
Column Total	3.00	47.00	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.30	0.00	0.30
Semantic -- N	0.04	0.07	0.11
Column Total	0.34	0.07	0.41

Contingency Table of Collocation Types for PO: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0	7	7
Phonology -- N	3	40	43
Column Total	3	47	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.42	6.58	7
Phonology -- N	2.58	40.42	43
Column Total	3.00	47.00	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	2.02	0.00	2.02
Phonology -- N	0.00	0.02	0.02
Column Total	2.02	0.02	2.04

Contingency Table of Collocation Types for PO: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	4	3	7
Phonology -- N	35	8	43
Column Total	39	11	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	5.46	1.54	7
Phonology -- N	33.54	9.46	43
Column Total	39.00	11.00	50

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.70	0.60	1.30
Phonology -- N	0.03	0.40	0.43
Column Total	0.73	1.00	1.73

Contingency Table of Collocation Types for PO: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	13	13
Grammar -- N	3	34	37
Column Total	3	47	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0.78	12.22	13
Grammar -- N	2.22	34.78	37
Column Total	3.00	47.00	50

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.11	0.00	2.11
Grammar -- N	0.03	0.05	0.08
Column Total	2.14	0.05	2.19

Contingency Table of Collocation Types for PO: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	4	9	13
Grammar -- N	35	2	37
Column Total	39	11	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	10.14	2.86	13
Grammar -- N	28.86	8.14	37
Column Total	39.00	11.00	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	4.35	11.12	15.47
Grammar -- N	1.10	5.42	6.52
Column Total	5.45	16.54	21.99

Contingency Table of Collocation Types for PO: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	3	10	13
Grammar -- N	4	33	37
Column Total	7	43	50

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	1.82	11.18	13
Grammar -- N	5.18	31.82	37
Column Total	7.00	43.00	50

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.25	0.26	0.51
Grammar -- N	0.54	0.01	0.55
Column Total	0.79	0.27	1.06

Contingency Table of Collocation Types for WI: Semantic vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	3	21	24
Semantic -- N	0	6	6
Column Total	3	27	30

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	2.4	21.6	24
Semantic -- N	0.6	5.4	6
Column Total	3.0	27.0	30

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Semantic -- Y	0.00	0.06	0.06
Semantic -- N	2.02	0.00	2.02
Column Total	2.02	0.06	2.08

Contingency Table of Collocation Types for WI: Phonology vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	1	6	7
Phonology -- N	2	21	23
Column Total	3	27	30

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.7	6.3	7
Phonology -- N	2.3	20.7	23
Column Total	3.0	27.0	30

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Phonology -- Y	0.06	0.10	0.16
Phonology -- N	0.28	0.00	0.28
Column Total	0.34	0.10	0.44

Contingency Table of Collocation Types for WI: Phonology vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	4	3	7
Phonology -- N	20	3	23
Column Total	24	6	30

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	5.6	1.4	7
Phonology -- N	18.4	4.6	23
Column Total	24.0	6.0	30

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Semantic -- Y	Semantic -- N	Row Total
Phonology -- Y	0.79	0.86	1.65
Phonology -- N	0.06	0.96	1.02
Column Total	0.85	1.82	2.67

Contingency Table of Collocation Types for WI: Grammar vs. Lexical Transfer

(a) Observed Frequencies

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0	9	9
Grammar -- N	3	18	21
Column Total	3	27	30

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	0.9	8.1	9
Grammar -- N	2.1	18.9	21
Column Total	3.0	27.0	30

(c) Deviances: [(observed-expected)-0.5]² ÷ expected

	Lexical Transfer -- Y	Lexical Transfer -- N	Row Total
Grammar -- Y	2.18	0.02	2.20
Grammar -- N	0.08	0.10	0.18
Column Total	2.26	0.12	2.38

Contingency Table of Collocation Types for WI: Grammar vs. Semantic

(a) Observed Frequencies

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	3	6	9
Grammar -- N	21	0	21
Column Total	24	6	30

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	7.2	1.8	9
Grammar -- N	16.8	4.2	21
Column Total	24.0	6.0	30

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Semantic -- Y	Semantic -- N	Row Total
Grammar -- Y	3.07	7.60	10.67
Grammar -- N	0.81	5.26	6.07
Column Total	3.88	12.86	16.74

Contingency Table of Collocation Types for WI: Grammar vs. Phonology

(a) Observed Frequencies

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	3	6	9
Grammar -- N	4	17	21
Column Total	7	23	30

(b) Expected Frequencies: (row total) x (column total) ÷ (grand total)

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	2.1	6.9	9
Grammar -- N	4.9	16.1	21
Column Total	7.0	23.0	30

(c) Deviances: $[(\text{observed}-\text{expected})-0.5]^2 \div \text{expected}$

	Phonology -- Y	Phonology -- N	Row Total
Grammar -- Y	0.08	0.28	0.36
Grammar -- N	0.40	0.01	0.41
Column Total	0.48	0.29	0.77

REFERENCES

- Aarts, Jan and Willem Meijs (eds.) 1990. *Theory and practice in corpus linguistics*. Amsterdam/Atlanta: Rodopi.
- _____, Pieter de Haan and Nelleke Oostdijk (eds.) 1993. *English language corpora: Design, analysis, and exploitation*. Amsterdam/Atlanta: Rodopi.
- Aijmer, Karin and Begnt Altenberg (eds.) 1991. *English corpus linguistics: Studies in honour of Jan Svartvik*. London/New York: Longman.
- Alexander, Richard J. 1984. Fixed expressions in English: Reference books and the teacher. *ELT Journal* 38:2.127-34.
- _____. 1987. Problems in understanding and teaching idiomaticity in English. *Anglistik & Englischunterricht* 32:105-22.
- Alverson, Hoyt. 1994. *Semantics and experience: Universal metaphors of time in English, Mandarin, Hindi, and Sesotho*. Baltimore and London: Johns Hopkins University Press.
- Atkins, Sue, Jeremy Clear and Nicholas Ostler. 1992. Corpus design criteria. *Literary and Linguistic Computing* 7:1.1-16.
- Bahns, Jens. 1993. Lexical collocations: A contrastive view. *ELT Journal* 47:1.56-63.
- _____. and Moira Eldaw. 1993. Should we teach EFL students collocations? *System* 21:1.101-14.
- Baker, Mona, Gill Francis and Elena Tognini-Bonelli (eds.) 1994. *Text and technology: In honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins.
- Beaugrande, Robert de. 1993. 'Register' in discourse studies: A concept in search of a theory. *Register analysis theory and practice*, ed. by M. Ghadessy, 7-25. London/New York: Pinter Publishers.
- Benson, Morton. 1985. Collocations and idioms. *Dictionaries, lexicography and language learning (ELT Documents 120)*, ed. by R. Ilson, 61-68. Oxford: Pergamon Press.
- _____. 1989. The collocational dictionary and the advanced learner. *Learner's dictionaries: State of the art (Anthology series 23)*, ed. by M.L. Tickoo, 84-93. Singapore: SEAMEO Regional Language Centre.

- _____. 1989. A collocational dictionary of Russian. *Slavic and East European Journal* 33:4.593-608.
- _____, Evelyn Benson, and Robert Ilson (eds.) 1986. *The BBI combinatory dictionary of English: A guide to word combinations*. Amsterdam/Philadelphia: John Benjamins.
- Berry-Rogghe, Godelieve L.M. 1973. The computation of collocations and their relevance in lexical studies. *The computer and literary studies*, ed. by A.J. Aitken, R.W. Bailey, and N. Hamilton-Smith, 103-112. Edinburgh: Edinburgh University Press.
- Bhatia, Vijay K. 1991. A genre-based approach to ESP materials. *World Englishes* 10:2.153-66.
- _____. 1993. *Analysing genre: Language use in professional settings*. London/New York: Longman.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- _____, Susan Conrad, and Randi Reppen. 1994. Corpus-based approaches to issues in applied linguistics. *Applied Linguistics* 15:2.169-89.
- Biria, Reza and M. Hassan Tahriian. 1994. The methodology factor in teaching ESP. *English for Specific Purposes* 13.93-101.
- Blocksmas, Laura and Mark van der Kraan. 1992. Collocations and categorial grammar. *Computational linguistics in the Netherlands: Papers from the second CLIN meeting (OTS Occasional Publication)*, ed. by J. van Eijck and W. Viol, 12-25. Utrecht: Onderzoeksinstuut voor Taal en Sproak.
- Bolinger, Dwight. 1975. *Aspects of language* (second edition). New York/Chicago/San Francisco/Atlanta: Harcourt Brace Jovanovich, Inc.
- Bowen, Donald J., Harold Madsen, and Ann Hilferty. 1985. *TESOL techniques and procedures*. Cambridge, MA: Newbury House Publishers.
- Briscoe, Ted. 1990. English noun phrases are regular: A reply to professor Sampson. *Theory and practice in corpus linguistics*, ed. by J. Aarts and W. Meijs, 45-60. Amsterdam/Atlanta: Rodopi.
- Brown, Dorothy F. 1974. Advanced vocabulary teaching: The problem of collocation. *RELJ Journal* 5:2.1-11.
- Brumfit, C.J. (ed.) 1984. *Common ground: Shared interest in ESP and communication studies*. Oxford: The British Council/Pergamon Press.
- Butler, Christopher S. (ed.) 1992. *Computers and written texts*. Cambridge, MA: Basil Blackwell.

- Carter, Ronald. 1987. *Vocabulary: Applied linguistic perspectives*. London: Allen & Unwin.
- _____ and Michael McCarthy, eds. 1988. *Vocabulary and language teaching*. London/New York: Longman.
- Celce-Murcia, Marianne and Diane Larsen-Freeman. 1983. *The grammar book*. Cambridge, MA: Newbury House Publishers.
- Chafe, Wallace. 1992. The importance of corpus linguistics to understanding the nature of language. *Directions in corpus linguistics: Proceedings of Nobel symposium 82 (Trends in Linguistics: Studies and Monographs 65)*, ed. by J. Svartvik, 79-97. Berlin/New York: Mouton de Gruyter.
- Channell, Joanna. 1981. Applying semantic theory to vocabulary teaching. *ELT Journal* 35.115-22.
- Chuang, Yuangshan. 1993. *A quantitative corpus analysis of word frequency and part of speech in the English textbooks used in senior high schools in Taiwan*. Urbana, IL: University of Illinois dissertation.
- Church, Kenneth and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16:1.22-29.
- _____, William Gale, Patrick Hanks and Donald Hindle. 1991. Using statistics in lexical analysis. *Lexical acquisition: Exploiting on-line resources to build a lexicon*, ed. by U. Zernik, 115-64. Hillsdale, NJ/London: Lawrence Earlbaum.
- Clear, Jeremy. 1992. Corpus sampling. *New directions in English language corpora*, ed. by G. Leitner, 21-31. New York: Mouton de Gruyter.
- _____. 1994. From Firth principles: Computational tools for the study of collocation. *Text and technology: In honour of John Sinclair*, ed. by M. Baker et al., 271-92. Amsterdam/Philadelphia: John Benjamins.
- Corder, S. Pit. 1981. *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Cowan, Laura. 1989. *Towards a definition of collocation*. Montreal: Concordia University MA Thesis.
- Cowie, Anthony P. 1978. The place of illustrative material and collocations in the design of a learner's dictionary. In honour of A.S. Hornby, ed. by P. Strevens, 127-39. Oxford: Oxford University Press.
- _____. 1991. The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics* 11:3.223-235.
- Crabbe, David. 1988. Design and description in ESP. *ESP: State of the art*, ed. by M.L. Tickoo, 53-57. Singapore: SEAMEO Regional Language Centre.

- Cross, Marilyn. 1993. Collocation in computer modelling of lexis as most delicate grammar. Register analysis theory and practice, ed. by M. Ghadessy, 196-220. London/New York: Pinter Publishers.
- Cruse, D.A. 1986. Lexical Semantics. Cambridge: Cambridge University Press.
- Dechert, Hans W. and Paul Lennon. 1989. Collocational blends of advanced second language learners: A preliminary analysis. Contrastive pragmatics, ed. by W. Olesky, 131-68. Amsterdam/Philadelphia: John Benjamins.
- Delacourt, Christian. 1992. About the statistical analysis of co-occurrence. Computers and the Humanities 26.21-29.
- Dudley-Evans, Tony. 1988. Recent developments in ESP: The trend to greater specialization. ESP: State of the art, ed. by M.L. Tickoo, 27-32. Singapore: SEAMEO Regional Language Centre.
- _____ and Willie Henderson (eds.) 1990. The language of economics: The analysis of economics discourse (ELT Documents 134). London: MacMillan Publishers, Ltd.
- Educational Testing Service. 1994. The official guide for GMAT review. Princeton, NJ: Educational Testing Service.
- Emery, P. G. 1988. Collocation - a problem in Arabic/English translation? *Quinquere* 11:2.178-184.
- Engwall, Gunnel. 1994. Not chance but choice: Criteria in corpus creation. Computational approaches to the lexicon, ed. by B.T.S. Atkins and A. Zampolli, 49-82. Oxford: Oxford University Press.
- Evens, Martha W., ed. 1988. Relational models of the lexicon. Cambridge: Cambridge University Press.
- Fang, Cheng-yu. 1993. Building a corpus of the English of computer science. English language corpora: Design, analysis, and exploitation, ed. by J. Aarts et al., 73-78. Amsterdam/Atlanta: Rodopi.
- Fanning, Paul. 1993. Broadening the ESP umbrella. *English for Specific Purposes* 12.159-70.
- Farghal, Mohammed and Hussein Obiedat. 1995. Collocations: A neglected variable in EFL. *IRAL* 33:4.315-330.
- FileMaker Pro, version 3.0. 1995. Computer software. Macintosh. The Claris Corporation.
- Firth, J.R. 1957. Modes of meaning. *Papers in Linguistics 1934-1951*, 190-215. London: Oxford University Press.
- Flower, John and Ron Martínez. 1995. American business vocabulary. Hove, England: Language Teaching Publications.

- Flowerdew, John. 1990. English for specific purposes—A selective review of the literature. *ELT Journal* 44:4.326-37.
- _____. 1993. An educational, or process, approach to the teaching of professional genres. *ELT Journal* 47:4.305-16.
- FolioVIEWS, version 3.1. 1994. Computer software. Macintosh. Provo, UT: Folio Corporation.
- Fontenelle, Thierry. 1992. Collocation acquisition from a corpus or from a dictionary: A comparison. *EURALEX '92 Proceedings*, ed. by Hannu Tommola et al., 221-28. Tampere, Finland: Tampereen Yliopisto.
- _____. 1994. What on earth are collocations? *English Today* 40 10:4.42-48.
- Francis, Gill and John Sinclair. 1994. I bet he drinks 'Carling Black Label': A riposte to Owen on corpus grammar. *Applied Linguistics* 15:2.190-200.
- Francis, W. Nelson and Henry Kucera. 1982. *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Frawley, W. 1988. Relational models and metascience. *Relational models of the lexicon*, ed. by M. Evens, 335-72. Cambridge: Cambridge University Press.
- Garside, Roger, Geoffrey Leech and Geoffrey Sampson (eds.) 1987. *The computational analysis of English: A corpus-based approach*. London/New York: Longman.
- Ghadessy, Mohsen. 1993. On the nature of written business communication. *Register analysis theory and practice*, ed. by M. Ghadessy, 149-164. London/New York: Pinter Publishers.
- Gitsaki, Christina. 1996. *The development of ESL collocational knowledge*. Queensland: University of Queensland dissertation.
<<http://www.cltr.uq.oz.au:8000/users/chrisitna.gitsaki/thesis>>
- Granger, Sylviane. 1993. *International corpus of learner English. English language corpora: Design, analysis, and exploitation*, ed. by J. Aarts et al., 57-71. Amsterdam/Atlanta: Rodopi.
- _____. 1994. The learner corpus: A revolution in applied linguistics. *English Today* 39:10.25-32.
- Grishman, Ralph and Richard Kittredge (eds.) 1986. *Analyzing language in restricted domains: Sublanguage description and processing*. Hillsdale, NJ: Lawrence Erlbaum.
- Haan, Pieter de. 1992. The optimum corpus sample size? *New directions in English language corpora*, ed. by G. Leitner, 3-19. New York: Mouton de Gruyter.

- Halliday, M.A.K., Angus McIntosh and Peter Strevens. 1964. *The linguistic sciences and language teaching*. Bloomington: Indiana University Press.
- _____ and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Harris, Catherine L. 1995. Exploring the continuum of unit size in word identification. WWW document
<<http://web.bu.edu/PSYCH/faculty/charris/papers/cogsci.paper/cogsci.html>>.
- Hatch, Evelyn and Cheryl Brown. 1995. *Vocabulary, semantics, and language education*. Cambridge: Cambridge University Press.
- Herbst, Thomas. 1996. What are collocations: Sandy beaches or false teeth? *English Studies* 77:4.379-393.
- Hoey, Michael. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.
- Hussein, Riyad Fayez. 1990. Collocations: The missing link in vocabulary acquisition amongst EFL learners. *Papers and Studies in Contrastive Linguistics* 26:123-36.
- Hutchinson, Tom and Alan Waters. 1987. *English for specific purposes: A learning-centred approach*. Cambridge: Cambridge University Press.
- Illson, Robert (ed.) 1985. *Dictionaries, lexicography and language learning (ELT Documents 120)*. Oxford: Pergamon Press.
- Johansson, Stig. 1995. ICAME—Quo vadis? Reflections on the use of computer corpora in linguistics. *Computers and the Humanities* 28.243-52.
- _____ and Anna-Brita Stenström (eds.) 1991. *English computer corpora: Selected papers and research guide*. Berlin/New York: Mouton de Gruyter.
- Johns, Ann M. 1980. Cohesion in written business discourse: Some contrasts. *The ESP Journal* 1:1.35-44.
- _____ and Tony Dudley-Evans. 1991. English for specific purposes: International in scope, specific in purpose. *TESOL Quarterly* 25:2.297-314.
- Jones, Gary M. 1990. ESP textbooks: Do they really exist? *English for Specific Purposes* 9.89-93.
- Kennedy, Graeme. Collocations: Where grammar and vocabulary teaching meet (ED 366 196). *Language teaching methodologies for the nineties*, ed. by A. Sarinee, 215-29. Singapore: SEAMEO Regional Language Centre.
- _____. 1992. Preferred ways of putting things with implications for language teaching. *Directions in corpus linguistics*, ed. by J. Svartvik, 335-73. Berlin/New York: Mouton de Gruyter.
- Khiralla, Mohammed Nagy A.K. 1990. The development of a computer-assisted maritime English language learning programme: Describing, testing and teaching some

stylistic features of international maritime conventions. Exeter, UK: University of Exeter dissertation.

- Kjellmer, Göran. 1987. Aspects of English collocations. *Corpus linguistics and beyond*, ed. by W. Meijs, 133-40. Amsterdam/Atlanta: Rodopi.
- _____. 1990. Patterns of collocability. *Theory and practice in corpus linguistics*, ed. by J. Aarts and W. Meijs, 163-78. Amsterdam/Atlanta: Rodopi.
- _____. 1992. Grammatical or nativelike? *New directions in English language corpora*, ed. by G. Leitner, 329-44. New York: Mouton de Gruyter.
- Knowles, Frank. 1993. Dictionaries for advanced learners and users of foreign languages. *Verbatim: The Language Quarterly* 19:3.13-16.
- Korosadowicz-Struzynska, Maria. 1980. Word collocations in FL vocabulary instruction. *Studia Anglica Posnaniensia* 12.109-120.
- Kroll, Barbara (ed.) 1990. *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press.
- Kytö, Merja, Ossi Ihalainen and Matti Rissanen (eds.) 1988. *Corpus linguistics, hard and soft (Proceedings of the 8th International Conference on English Language Research on Computerized Corpora)*. Amsterdam: Rodopi.
- Lai, Jennifer C. 1996. A beginner's guide to collocations (Web page). WWW address: <http://www.ed.uiuc.edu/students/jc-lai/Fall95/collocations.html> (viewed 1-13-96).
- Laufer, Batia. 1994. The lexical profile of second language writing: Does it change over time? *RELC Journal* 25:2.21-33.
- Leech, Geoffrey. 1993. Corpus annotation schemes. *Literary and Linguistic Computing* 8:4.275-81.
- _____. and Steven Fligelstone. 1992. Computers and corpus analysis. *Computers and written texts*, ed. by C. Butler, 115-40.
- Leitner, Gerhard. 1992. International corpus of English: Corpus design—problems and suggested solutions. *New directions in English language corpora*, ed. by G. Leitner, 33-64. New York: Mouton de Gruyter.
- _____. (ed.) 1992. *New directions in English language corpora*. New York: Mouton de Gruyter.
- Lewis, Michael. 1993. *The lexical approach*. Hove, England: Language Teaching Publications.
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

- Li, Charles N. and Sandra Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley and Los Angeles: University of California Press.
- Ljung, Magnus. 1991. Swedish TEFL meets reality. *English computer corpora: Selected papers and research guide*, ed. by S. Johansson and A. Stenström, 245-56. Berlin/New York: Mouton de Gruyter.
- Lombard, Robin. 1995. Corpus analysis and advanced learner lexis. Paper presented at the Linguistic Association of the Southwest 25th Annual Conference, October 1995.
- Long, Mike. 1990. Collocation and lexical development for language learners. *PASAA* 20:1.35-40.
- Love, Alison. 1993. Lexico-grammatical features of geology textbooks: Process and product revisited. *English for Specific Purposes* 12.197-218.
- Mackin, Ronald. 1978. On collocations: Words shall be known by the company they keep. In honour of A.S. Hornby, ed. by P. Strevens, 149-65. Oxford: Oxford University Press.
- Mair, Christian. 1988. In defense of *the fact that*: A corpus-based study of current British Usage. *Journal of English Linguistics* 21:1.59-71.
- _____. 1991. Quantitative or qualitative corpus analysis? Infinitival complement clauses in the Survey of English Usage corpus. *English computer corpora: Selected papers and research guide*, ed. by S. Johansson and A. Stenström, 67-80. Berlin/New York: Mouton de Gruyter.
- Malmkjær, Kirsten. 1994. Who can make *nice* a better word than *pretty*? Collocation, translation, and psycholinguistics. *Text and technology: In honour of John Sinclair*, ed. by M. Baker et al., 211-32. Amsterdam/Philadelphia: John Benjamins.
- Martin, Marilyn. 1984. Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal* 68:2.130-37.
- Marton, Waldemar. 1977. Foreign vocabulary learning as problem no.1 of language teaching at the advanced level. *Interlanguage Studies Bulletin* 2.33-57.
- McCarthy, Michael. 1990. *Vocabulary*. Oxford: Oxford University Press.
- McIntosh, Angus. 1961. Patterns and ranges. *Language* 37:3.325-37.
- Meijs, Willem (ed.) 1987. *Corpus linguistics and beyond*. Amsterdam: Rodopi.
- _____. 1992. Computers and dictionaries. *Computers and written texts*, ed. by C. Butler, 141-65. Cambridge, MA: Basil Blackwell.

- Mel'čuk, Igor A. 1986. Semantic bases of linguistic description (meaning-text linguistic theory). The twelfth LACUS forum, ed. by M. C. Marino and L. A. Pérez, 41-84. Lake Bluff, IL: Linguistic Association of Canada and the United States.
- _____ and Alexander Zholkovsky. 1988. The explanatory combinatorial dictionary. Relational models of the lexicon, ed. by M. Evens, 41-74. Cambridge: Cambridge University Press.
- Mitchell, T. F. 1971. Linguistic 'goings on': Collocations and other lexical matters arising on the syntagmatic record. *Archivum Linguisticum* 2.35-69.
- Müller, Ernst-August. 1981. Towards a stratificational description of collocations. The seventh LACUS forum, ed. by J. E. Copeland and P. W. Davis, 175-187. Columbia, SC: Hornbeam Press.
- Nagy, William. 1978. Some non-idiom larger-than-word units in the lexicon. Papers from the parasession on the lexicon, ed. by D. Farkas, W. M. Jacobsen, and K. W. Todrys, 289-300. Chicago: Chicago Linguistic Society.
- Nation, I. S. P. 1990. Teaching and learning vocabulary. Boston, MA: Heinle & Heinle.
- Nattinger, James. 1988. Some current trends in vocabulary teaching. Vocabulary and language teaching, ed. by R. Carter and M. McCarthy, 62-82. London/New York: Longman.
- _____. and Jeannette S. DeCarrico. 1992. Lexical phrases and language teaching. New York: Oxford University Press.
- Newmark, Peter. 1973. Twenty-three restricted rules of translation. *The Incorporated Linguist* 12:1.9-15.
- Odlin, Terence. 1989. Language transfer. New York: Cambridge.
- Oostdijk, Nelleke. 1991. Corpus linguistics and the automatic analysis of English (Language and Computers: Studies in Practical Linguistics No. 6). Amsterdam/Atlanta: Rodopi.
- Owen, Charles. 1993. Corpus-based grammar and the Heineken effect: Lexico-grammatical description for language learners. *Applied Linguistics* 14:2.167-87.
- Palmer, H.E. 1933. Second interim report on English collocations. Tokyo: Institute for Research in English Teaching.
- Pienemann, Manfred. 1992. COALA—A computational system for interlanguage analysis. *Second Language Research* 8:1.59-92.
- Pohlsward, Ruja. 1993. The English language needs of Thai computing professionals. *RELC Journal* 24:1.86-108.

- ProIcon programming language for Apple Macintosh computers, version 2.0. 1990. Salida, CO and Tucson, AZ: The ProIcon Group.
- Renouf, Antoinette. 1988. Coding metalanguage: Issues raised in the creation and processing of specialised corpora. *Corpus linguistics, hard and soft* (Proceedings of the 8th International Conference on English Language Research on Computerized Corpora), ed. by M. Kytö et al., 197-206. Amsterdam: Rodopi.
- Robinson, Pauline. 1991. *ESP today: A practitioner's guide*. New York/London: Prentice Hall.
- Roos, Eckhard. 1976. Contrastive collocational analysis. *Papers and studies in contrastive linguistics, volume five*, ed. by J. Fisiak, 65-75. Poznan: Adam Mickiewicz University.
- Sampson, Geoffrey. 1987. Evidence against the "grammatical"/"ungrammatical" distinction. *Corpus linguistics and beyond*, ed. by W. Meijs, 219-26. Amsterdam: Rodopi.
- Sardinha, A. P. Berber. 1995. *Lexical co-occurrence frameworks: A preliminary investigation into Business English phraseology*. Unpublished ms.
- _____. 1996. Corpus choices in a short journalistic text. *The ESPecialist* 16:1.1-19.
- Savage, William and Rick Whisenand. 1993. Logbooks and language learning objectives in an intensive ESP workshop. *TESOL Quarterly* 27:4.741-46.
- Schenk, André Yvon. 1994. *Idioms and collocations in compositional grammars* (OTS Dissertation Series). Utrecht: Onderzoeksinstituut voor Taal en Spraak.
- Schmied, Josef. 1993. Qualitative and quantitative research approaches to English relative constructions. *Corpus-based computational linguistics*, ed. by C. Souter and E. Atwell, 85-96. Amsterdam/Atlanta: Rodopi.
- Seal, Bernard David. 1981. *In search of significant collocations*. Los Angeles: University of California MA thesis.
- Sinclair, John M. 1985. *Lexicographic evidence. Dictionaries, lexicography and language learning* (ELT Documents 120), ed. by R. Ilson, 81-94. Oxford: Pergamon Press.
- _____. 1987a. Collocation: A progress report. *Language topics: Essays in honour of Michael Halliday, vol. 2*, ed. by R. Steele and T. Threadgold, 319-31. Amsterdam/Philadelphia: John Benjamins.
- _____. (ed.) 1987b. *Collins COBUILD English language dictionary*. London: Collins.
- _____. (ed.) 1987c. *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. London/Glasgow: Collins ELT.

- _____. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Smadja, Frank. 1989. Lexical co-occurrence: The missing link. *Literary and Linguistic Computing* 4:3.163-68.
- _____. 1993a. Retrieving collocations from text: XTRACT. *Computational Linguistics* 19:1.143-77.
- _____. 1993b. XTRACT: An overview. *Computers and the Humanities* 26.399-413.
- Stevens, Vance. 1991. Classroom concordancing: Vocabulary materials derived from relevant, authentic text. *English for Specific Purposes* 10:1.35-46.
- Stevens, Peter. 1988. ESP after 20 years: A re-appraisal. *ESP: State of the art*, ed. by M.L. Tickoo, 1-13. Singapore: SEAMEO Regional Language Centre.
- Stubbs, Michael. 1993. British traditions in text analysis. From Firth to Sinclair. *Text and technology: In honour of John Sinclair*, ed. by M. Baker et al., 1-33. Amsterdam/Philadelphia: John Benjamins.
- _____. 1994. Grammar, text, and ideology: Computer-assisted methods in the linguistics of representation. *Applied Linguistics* 15:2. 201-23.
- _____. 1995. Collocations and semantic profiles. On the cause of the trouble with quantitative studies. *Functions of Language* 2:1.23-55.
- _____. 1995. Collocations and cultural connotations of common words. *Linguistics and Education* 7:4.379-390.
- Summers, Della. 1993. Longman/Lancaster English language corpus—Criteria and design. *International Journal of Lexicography* 6:3.181-208.
- Sutarsyah, Cucu, Paul Nation and Graeme Kennedy. 1994. How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal* 25:2.34-50.
- Svartvik, Jan (ed.) 1992. *Directions in corpus linguistics: Proceedings of Nobel symposium 82 (Trends in Linguistics: Studies and Monographs 65)*. Berlin/New York: Mouton de Gruyter.
- Swales, John M. 1985. *Episodes in ESP*. New York: Pergamon Institute of English.
- _____. 1988a. ESP and applied linguistics: Hopes for a brave new world. *ESP: State of the art*, ed. by M.L. Tickoo, 14-20. Singapore: SEAMEO Regional Language Centre.
- _____. 1988b. Communicative language teaching in ESP contexts. *Annual Review of Applied Linguistics* 8.48-57.
- _____. 1990. *Genre analysis: English in academic and research settings*. New York: Cambridge University Press.

- Swan, Michael and Bernard Smith (eds.) 1987. *Learner English: A teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.
- Taylor, Linda. 1992. *Vocabulary in action*. London: Prentice Hall.
- Tesitelova, Marie. 1992. *Quantitative linguistics*. Amsterdam/Philadelphia: John Benjamins.
- Thoiron, Philippe. 1994. Topography of terms and structure of techno-scientific texts. *Research in humanities computing 2*, ed. S. Hockey and N. Ide, 23-33. Oxford: Clarendon Press.
- Thomas, Helen. 1984. Developing the stylistic and lexical awareness of advanced students. *ELT Journal* 38:3.187-191.
- Thomson, Norman. 1989. How to read articles which depend on statistics. *Literary and Linguistic Computing* 4:1.6-11.
- Tickoo, M.L. (ed.) 1988. *ESP: State of the art*. Singapore: SEAMEO Regional Language Centre.
- Tuck, Allene (ed.) 1993. *Oxford dictionary of business English*. Oxford: Oxford University Press.
- van Dijk, Teun A. 1987. Discourse analysis and computer analysis. A few notes for discussion. *Corpus linguistics and beyond*, ed. by W. Meijs, 269-83. Amsterdam: Rodopi.
- Vikner, Carl. 1991. Problems in the use of text corpora in linguistic research. *LSP: Nine studies on language for special purposes (Copenhagen Studies in Language 14)*, ed. by N. Davidsen-Nielsen, 15-23. Copenhagen: Forlag.
- Weinert, Regina. 1995. The role of formulaic language in second language acquisition: A review. *Applied Linguistics* 16:2.180-205.
- Widdowson, H.G. 1984. *Learning purpose and language use*. Oxford: Oxford University Press.
- Williams, Marion. 1988. Language taught for meetings and language used in meetings: Is there anything is common? *Applied Linguistics* 9:1.45-58.
- Woods, Anthony, Paul Fletcher and Arthur Hughes. 1986. *Statistics in language studies*. Cambridge: Cambridge University Press.
- Wu, Hui Hua Marian. 1992. Towards a contextual lexico-grammar: An application of concordance analysis in EST teaching. *RELC Journal* 23:2.18-34.
- Yip, Virginia. 1995. *Interlanguage and learnability: From Chinese to English*. Amsterdam/Philadelphia: John Benjamins.

Zernik, Uri (ed.) 1991. *Lexical acquisition: Exploiting on-line resources to build a lexicon*. Hillsdale, NJ: Lawrence Erlbaum.

Zughoul, Muhammad Raji. 1991. *Lexical choice: Towards writing problematic word lists*. IRAL 29:1.45-60.