

THE EFFECTS OF EXPLICIT INSTRUCTION OF FORMULAIC SEQUENCES  
ON SECOND-LANGUAGE WRITERS

by

Jelena Čolović-Marković

A dissertation submitted to the faculty of  
The University of Utah  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

The University of Utah

December 2012

Copyright © Jelena Čolović-Marković 2012

All Rights Reserved

**The University of Utah Graduate School**

**STATEMENT OF DISSERTATION APPROVAL**

The dissertation of Jelena Čolović-Marković  
has been approved by the following supervisory committee members:

<u>MaryAnn Christison</u>	, Chair	<u>May 16, 2012</u> Date Approved
<u>Edward J. Rubin</u>	, Member	<u>May 15, 2012</u> Date Approved
<u>Rachel Hayes-Harb</u>	, Member	<u>May 15, 2012</u> Date Approved
<u>C. Ray Graham</u>	, Member	<u>May 15, 2012</u> Date Approved
<u>Mark Davies</u>	, Member	<u>May 15, 2012</u> Date Approved

and by Edward J. Rubin, Chair of  
the Department of Linguistics

and by Charles A. Wight, Dean of The Graduate School.

## **ABSTRACT**

The present study investigated the effects of the explicit teaching of formulaic sequences (i.e., academic and topic-induced) on L2 writing. The research examined separately the effects of the treatment on the students' abilities to produce the target formulaic sequences in controlled (i.e., C-tests) and uncontrolled situations (i.e., essays), and to produce better quality essays. The study, through posttreatment interviews, also attempted to glean insights into the approaches L2 writers use for the production of the target formulaic sequences.

The study found that the students in the treatment condition performed at a significantly higher level than the students in the control condition on measures of the production of academic formulaic sequences in a controlled situation and the production of topic-induced formulaic sequences in controlled and uncontrolled situations, but there were no significant differences between the groups on the measures of students' abilities to produce academic formulaic sequences in an uncontrolled situation and to produce better quality writing.

The results of the study suggest that the explicit instruction facilitated learning of the target formulaic sequences when the learning is measured by a test. Most importantly, the study found that explicit instruction helped students become familiar enough with the formulaic sequences to recognize their usefulness and employ them in their essays. The results indicated that the instructional approach helped low performing writers produce

essays of better quality and high performing writers increase the awareness of the frequencies and functions of the formulaic sequencers in expert writers' texts.

The findings from the interviews indicated that the students' abilities to produce the academic and topic-induced formulaic sequences in their compositions may depend on the students' perceived need to use them in their writing, and that students' abilities to produce the academic formulaic sequences in essays may be influenced by the interaction of students' awareness of the frequency and functions of academic formulaic sequences and their motivation to sound academic in their writing.

The study results suggest that the instructional approach may be helpful for the students' learning of formulaic sequences for the purposes of writing and should be the focus of future experimental research.

## TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	vii
Chapters	Page
1 INTRODUCTION.....	1
Research Questions.....	6
Organization of the Dissertation.....	7
2 REVIEW OF THE LITERATURE.....	9
The Role of Vocabulary in L2 Writing.....	9
Vocabulary of Academic Prose.....	12
Formulaic Sequences: Forms, Functions and Terminology.....	13
Formulaic Sequences in L2 Writing: Benefits and Challenges.....	17
The Usage and Acquisition of Formulaic Sequences by L1 and L2 Writers.....	18
The L2 Writers' Use of Formulaic Sequences.....	18
L1 Learners' Use of Formulaic Sequences.....	26
L2 Learners' Acquisition of Formulaic Sequences.....	29
3 METHODOLOGY.....	41
Overview of the Research Design.....	42
Selection of Target Formulaic Sequences.....	44
Participants.....	52
Teacher.....	53
Instruments for Quantitative Data Elicitation and Evaluation.....	55
Description of the Treatment.....	62
Study Procedures for Quantitative Data Collection.....	68
Study Procedures for Quantitative Data Analysis.....	70
Study Procedures for Qualitative Data Elicitation.....	76
4 RESULTS.....	81
Descriptive Statistics.....	81
Differences between Control and Experimental Groups.....	84
Qualitative Analyses from the Follow-up Interviews.....	87

5 DISCUSSION .....	99
Discussion of the Results of Quantitative Data Analysis .....	99
The Effects of Explicit Instruction of the Academic Formulaic Sequences on L2 Writing .....	101
The Effects of explicit instruction of Topic-Induced Formulaic Sequences on L2 Writing .....	109
The Effects of Explicit Instruction of the Academic and Topic-induced Formulaic Sequences on the Quality of L2 Writing.....	112
Discussion of the Results of Qualitative Data Analysis .....	114
Implications for L2 Writing Pedagogy .....	117
6 CONCLUSION .....	123
Concluding Remarks.....	123
Study Limitations and Implications for Future Research .....	125
Appendices	
A: CORE ACADEMIC FORMULAS AFL FOUND IN THE READING MATERIALS OF AN ADVANCED WRITING CLASS ARRANGED BY PRAGMATIC FUNCTION .....	129
B: WRITTEN ACADEMIC FORMULAS AFL FOUND IN THE READING MATERIALS OF AN ADVANCED WRITING CLASS ARRANGED BY PRAGMATIC FUNCTION .....	134
C: TOPIC-INDUCED FORMULAIC SEQUENCES BY ALPHA LISTING BY THEME .....	138
D: BACKGROUND SURVEY .....	140
E: THE C-TEST ON PRODUCTION OF THE ACADEMIC FORMULAIC SEQUENCES.....	141
F: THE C-TEST ON PRODUCTION OF THE TOPIC-INDUCED FORMULAIC SEQUENCES.....	145
G: IN-CLASS ARGUMENTATIVE ESSAY .....	146
H: ESSAY RATING RUBRIC .....	147
I: INDEPENDENT SAMPLES TEST RESULTS .....	148
REFERENCES .....	149

## ACKNOWLEDGEMENTS

Writing this dissertation would not have been possible without the support of my committee members, family, and friends. First of all, I would like to express my gratitude to my dissertation chair, MaryAnn Christison, her constant support during the writing process of the dissertation, her detailed and insightful feedback on the individual chapters, and above all, her commitment to my professional development. I would also like to thank my committee members, Dr. Edward Rubin, Dr. Rachel Hayes-Harb, Dr. C. Ray Graham, and Dr. Mark Davies for their helpful feedback on my topic from a variety of perspectives and their overall support. I am also very grateful to Dr. Nick Ellis for offering his comments in the early stages of this research.

A special thanks goes to the course instructor, for agreeing to implement the pedagogical intervention in her class and for working closely with me to help me realize the research project as designed; to George Plautz, the Director of the Intensive English Institute at the University of Utah, for letting me conduct the research at this institution; to the study participants; and to my raters for investing their time, on more than one occasion, to assist me in data collection.

I owe my deepest gratitude to my family. I am very thankful to my mom Dragica for the many personal sacrifices she has made to see me finish, for her unflagging support throughout the years, and for caring for my family when I was writing. I am indebted to my husband, Sasa, who stood by my side over the years, for his emotional and



professional support, and for taking care of our children, Dunya, Kosta, Voyin, and Dimitrije when I was engrossed in the project. Thank you Dunya, Kosta, Voyin, and Dimitrije, for your patience while I was writing, for reminding me of a different world outside the academia, and for believing in me.

## CHAPTER 1

### INTRODUCTION

Academic prose is marked by formal vocabulary composed of a variety of words, generally abstract, long, and content specific (Corson, 1985; 1995; Coxhead, 2000, 2008). Words such as *preliminary* and *coincide* that appear often in academic writing (Coxhead, 2000) probably have an unlikely chance of occurring in casual conversations. Corson (1995) refers to this academic vocabulary with its restricted use as a “lexical bar,” an obstacle that students (including English-as-a-second-language (ESL) learners) who intend to become a part of the higher education, academic culture have to negotiate at some point in their studies. Academic prose, as evidenced by the recent research in corpus linguistics (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Cortes, 2006; Hyalnd, 2008), is also marked by multiword combinations, phrases, or formulaic sequences larger than single words and by their high frequency and great functional utility. *In the context of* or *from the point of view of*, are similar to *preliminary* and *coincide* in that they are unlikely candidates for everyday communication. The presence of these phrases seems to create an additional barrier that native and non-native English speakers alike need to transcend to become successful in the community of writers in academia.

L1 writers are expected to possess a strong vocabulary base and the knowledge of the restrictions of use imposed by the register. Second language (L2) writers, regardless of their language background, within the academic context are expected to be able to use effectively the vocabulary of academic prose--the vocabulary identified by formal lexis, which includes high density content vocabulary and frequent and functional multiword combinations (Hinkel, 2004; Jones & Haywood, 2004; Folse, 2008). In other words, they need to have a large repertoire of words, as well as a knowledge of how these words combine together in the target language and which combinations are characteristic in the register of academic prose. These seem to be high expectations for L2 writers, to say at least.

The research on L1 learners' use of multiword combinations shows that they diverge from expert writers in the discipline (Cortes, 2004). The research on L2 learners' use of formulaic sequences in academic writing conducted through quantitative examinations of overall frequency of use and qualitative examinations of formulaic sequences produced by non-native writers highlights the issues related to the use of the sequences in their writing. The problems span a range from an overall underuse of formulaic sequences (Howarth, 1998), a failure to employ sequences present in the texts of expert writers in academia (Hyland, 2008; Scott & Tribble, 2006), an overreliance on a limited number of multiword sequences (Granger, 1998), and the creation of nontarget constructions (Erman, 2009). These problems taken together contribute to L2 writers' texts sounding awkward and foreign (Granger, 1998). Despite an extensive experience with and long practice in writing in the target language, the issues with formulaic sequence persist in ESL learners' texts (Li & Schmitt, 2008; Scott & Tribble, 2006),

which suggests that mastering the skillful and expert-like use of formulaic sequences expected in academic prose poses a great challenge for L2 writer.

The research has recognized the fact that learners struggle with using the formulaic sequences in an expert-like manner in their writing, which is evident in Simpson-Vlach and Ellis (2010), a most recent attempt to inform ESL writing pedagogy of the useful formulaic sequences for ESL writers. The researchers created a list of frequent and pedagogically useful formulaic sequences entitled *Academic Formulas List* which, among lists that pertain to spoken academic language, contains formulaic sequences that are found in academic prose, have a degree of pedagogical utility as measured by an empirically derived measure, and include sequences consisting of words with strong levels of association. Such an empirically derived list seems to be a very useful addition to L2 writing pedagogy.

The research in content-based instruction (CBI) emphasizes that in order to discuss effectively a particular topic, the vocabulary necessitated by that topic (i.e., content-obligatory) needs to be employed (Murray & Christison, 2010). Similarly, in the area of collocational studies that focus on formulaic language in L2 writing, Erman (2009) argues that a topic necessitates the use of specific word combinations, if the text adequately examines the issues related to the topic. In an event that a text is lacking the collocations induced by a topic, it may appear impoverished in content, and therefore, ineffective in the treatment of the topic. The students in an ESL writing class may not be writing within a particular content area, but are asked to write on a variety of topics which, similar to CBI, can be dealt with effectively when appropriate vocabulary is used. It does not seem enough to know the meaning of individual words without knowing the

ways in which they combine with other words in context and being able to manipulate formulaic sequences related to the discussion topic of a paper.

There needs to be more research conducted to investigate the processes involved in learning formulaic sequences for the purposes of production in writing (Cortes, 2006; Jones & Haywood, 2004; Li & Schmitt, 2008; Schmitt, 2004). Due to the lack of research on L1 and L2 on learners' acquisition of formulaic sequences, the field has turned to the field of vocabulary acquisition under the assumption that the conditions (i.e., noticing, production, and generation) needed to facilitate the learning of a word (Nation, 2001) would apply to the learning of a formulaic sequence. The research that investigated the processes involved in the learning of formulaic sequences for the purposes of written production in an L2, a longitudinal case study (Li and Schmitt, 2008) and longitudinal studies of the effects of explicit instruction on L2 learners (Jones and Haywood, 2004; Schmitt, Dorneyei, Adolphs, & Durow, 2004), point out that the learning of formulaic sequences is incremental, similar to that of learning single vocabulary, suggesting that relying on the field of vocabulary acquisition research is a viable approach. To this end, few previous investigations on the effects of explicit instruction of formulaic sequences (Jones and Haywood, 2004; Schmitt et al., 2004), which to various extents followed the practices of effective vocabulary teaching, report gains in learners' production in a controlled situation (i.e., C-test). However, these studies suffer from several shortcomings that challenge their results and call for more research to be conducted.

To illustrate, Schmitt et al. (2004) did not control the input that the study participants were receiving and did not include a control group in their study; as a result, the researchers were not able to attribute the statistically-reliable gains in production of

formulaic sequences in a controlled situation to explicit instruction. There remained the question whether the learners' gains in production in a controlled situation were a result of a rich ESL environment in which the participants were immersed, the ESL language program they attended, the explicit instruction they received, or, perhaps, all of them combined. In addition, the study did not attempt to investigate whether the participants were able to produce the target formulaic sequences in an uncontrolled situation, thus leaving the question as to whether learners are able to produce the target formulaic sequences independently in their own writing unaddressed.

Although the study by Jones and Haywood (2004) addressed the question of the use of formulaic sequences in an uncontrolled situation, it also has several weaknesses, two of which seem to be most serious--the small number of participants (i.e., 10 in the treatment group) and the relatively short period of time (i.e., 2 weeks) devoted to explicit instruction of formulaic sequences before data were collected on a posttest of production (i.e., essay). The researchers observed marked gains in identification of formulaic sequences, small gains in controlled production (i.e., C-test), and no gains in uncontrolled production (i.e., essay) between the group that received and the group that did not receive explicit instruction.

The two studies call for more experimental research to be conducted over an extended period of time to investigate the effects of explicit instruction of formulaic sequences on L2 writing. The present study is a step in that direction. It builds on Jones and Haywood (2004) study by attempting to correct the study limitations, but it also considers the necessity for specific formulaic sequences to be used in an effective

discussion of a topic, a necessity that has been recognized in the CBI and L2 writing literature.

### **Research Questions**

The current study attempts to answer the overarching question:

What is the effect of explicit instruction of formulaic sequences on L2 writing?

In order to answer the overarching question and document the effects of explicit instruction on the level of knowledge of formulaic sequences the following sub-questions need to be answered:

1. Is there a significant difference in students' abilities to *produce*, in a controlled situation (i.e., C-test), formulaic sequences frequent in academic prose (i.e., selected AFL [Simpson-Vlach & Ellis, 2010]) between the students who receive explicit instruction and those who do not?
2. Is there a significant difference in students' abilities to *produce*, in a controlled situation (i.e., C-test), formulaic sequences in topic discussion between the students who receive explicit instruction and those who do not?
3. Is there a significant difference in students' abilities to *produce*, in an uncontrolled situation (i.e., an essay), formulaic sequences frequent in academic prose (i.e., selected AFL [Simpson-Vlach & Ellis, 2010]) between the students who receive explicit instruction and those who do not?
4. Is there a significant difference in students' abilities to *produce*, in an uncontrolled situation (i.e., an essay), formulaic sequences in topic discussion between the students who receive explicit instruction and those who do not?

5. Is there a significant difference in the overall quality of students' essays as measured by the trained essay raters' judgments between the students who receive explicit instruction and those who do not?

In addition to answering the aforementioned questions, the study attempts to glean insights into the strategies ESL learners use in producing the formulaic sequences in their writing. With this goal in mind, individual interviews will be conducted at the end of the treatment with a subset of students who will be selected based on their performance profiles on the production tasks (see Chapter 3).

### **Organization of the Dissertation**

The present study is organized into five chapters. Chapter 1 provides an introduction into the problem of L2 learners' use of formulaic sequences. It highlights the need for writing teachers' interventions and a better understanding of the effects of explicit instruction of formulaic sequences in L2 writing and presents the questions that lead the investigation. Chapter 2 offers a review of relevant literature that begins with an discussion of the importance vocabulary has for ESL writers in an academic context, the type of vocabulary that the ESL learners need in order to function in academia and the need for explicit instruction, the contributions that corpus linguistics investigations have had in this regard, the approaches to instruction of vocabulary, and the validity of the research questions (Bachman & Palmer, 2010) for the present study. Chapter 3 presents the methodology for the study starting with the study overview and description of the participants and moving on to the identification of the target formulaic sequences, instrument design, and concluding with data collection and the proposed data analysis



procedures. Chapter 4 offers the results of the investigation while Chapter 5 presents a discussion of the results. Chapter 6, the conclusion, outlines implications from the present study and offers directions for further research.

## **CHAPTER 2**

### **REVIEW OF THE LITERATURE**

The present chapter provides theoretical background that motivates an investigation of the effects of explicit instruction of formulaic sequences on L2 writing. With this goal as a guide, the chapter begins with a discussion of the role of vocabulary in the academic writing of ESL learners. It moves on to highlight the recent contributions of the research in corpus linguistics that reveal information about multiword combinations or formulaic sequences that prevail in academic prose and describe the forms and functions of the phrases in academic texts. Then, the chapter discusses the studies describing the actual use of formulaic sequences by L1 and L2 learners to point out the difficulties L2 learners have in using the sequences in a target-like manner in their writing. It also outlines the attempts that have been made to assist L2 learners in dealing with issues related to the use of formulaic sequences. Finally, the chapter discusses a previous study that the present study builds on, pointing out the places in which the two studies overlap but also depart.

#### **The Role of Vocabulary in L2 Writing**

Vocabulary plays a critical role in the successful academic writing of L2 learners (Coxhead, 2008; Coxhead & Byrd, 2007; Nation, 2001, 2005; Folse, 2008; Hinkel, 2004,

Jones & Haywood, 2004; Li & Schmitt, 2009, Raimes, 1985). With a good command of English words (i.e., a large repertoire of English words), ESL learners can perform important tasks that help them succeed in a university setting: they can comprehend the information in the sources to cite, summarize, and paraphrase accurately; they can express complex ideas using abstract language; and they can produce readable and coherent prose their readers expect (Folse, 2008; Hinkel, 2004). Empirical evidence suggests that vocabulary utilized in L2 students' writing may influence the overall quality of an essay (Barkaoui, 2010; Engber, 1995; Ferris, 1994; Harley & King, 1989; Linnarud, 1986; McClure, 1991; Santos, 1988; Song & Caruso, 1996) and that effective lexical choices are contributing factors in the quality of an ESL student's text (Engber, 1995; Harley & King, 1989; Linnarud, 1986; McClure, 1991).

Any deficiencies in vocabulary knowledge can prevent ESL learners from completing the aforementioned tasks, hindering their academic progress. The lack of vocabulary can present a major barrier in written communication in a university setting. It can cause students to make poor lexical choices and also lexical errors in their writing. These errors are considered the most serious errors in L2 students' writing (Santos, 1988) probably because of the effect they have on communication of meaning, or in Santos' words "[i]t is precisely with this type of error that language impinges directly on content; when a wrong word is used, the meaning is very likely to be obscured" (p. 48).

Vocabulary is frequently included as one of the components in the rubrics developed as writing assessment tools. For example, it is a part of the widely used ESL Composition Profile, an analytic scoring rubric developed by Jacobs, Harfield, Hughey, & Wormeth (1981). The rubric used in this study is constructed to consider five areas of

writing each independently contributing to the final score in the following manner: content 30, organization 20, vocabulary 20, language use 25, and mechanics 5 for a total of 100 points.

Similarly, vocabulary is regarded as a factor influencing the overall score an essay receives on high-stakes tests, one of which is the written portion of the Examination for the Certificate of Competency in English (ECCE<sup>1</sup>) published by Cambridge Michigan. The scoring rubric for this test is designed to measure four constructs, two of which directly refer to the role of vocabulary choices in the text: a) organization and connection of ideas and b) linguistic range and control (Ohlrogge, 2009, p. 379). To receive a high score on the former measure, a learner needs to employ various cohesive devices and on the second demonstrate strong language control of broad range of vocabulary.

The importance of vocabulary in academic writing is also seen from the perspective of ESL writers. To illustrate, Leki and Carson (1994) have conducted a survey of 128 ESL undergraduate students on their English for Academic Purposes (EAP) writing course to gather data on the student perceived effectiveness of the writing course. One of the survey questions is an open-ended question asking the students to report on what it is they would like to have learned better in their writing courses. Among five skill categories (i.e., language skills, task management strategies, rhetorical skills, thinking skills, and other), language skills were ranked first. More importantly, within the language skills category, it was vocabulary that the students reported they had needed the most.

---

<sup>1</sup> ECCE is English as a Foreign Language (EFL) certification test designed for the B2 Level of the Common European Framework of Reference administered twice a year in 25 countries in Europe, Asia, and Latin America. The test assesses four language skills: reading (grammar and vocabulary included), listening, speaking, and writing. The test results can be used for purposes related to personal, public, educational, and vocational needs.

In summary, the evidence coming from research on the factors contributing the ESL essay quality, assessment tools used in evaluation of ESL essays, and students' perceptions of what needs to be included in the ESL writing instruction, suggests that vocabulary is considered critical in writing in an academic setting by the expert readers, ESL instructors and evaluators, and ESL students alike. The evidence from the three sources suggests that ESL learners need instruction in the use of vocabulary that effective academic writing requires. The next section discusses the vocabulary found by corpus linguistics to be prevalent in academia and as such expected to be included in the ESL writing in an academic setting.

### **Vocabulary of Academic Prose**

Corpus linguistics research focuses on a systematic investigation of authentic texts, written or spoken, that are stored on the computer and available for qualitative and quantitative analyses that most frequently utilizes computer software. These corpora have been instrumental in describing the make-up and identifying the distinguishing features of academic discourse. Recent vocabulary research in corpus linguistics reveals that vocabulary used in academia is often made up of multiword combinations, that is, phrases larger than single words (Biber & Barbieri, 2007; Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Cortes, 2002, 2004, 2006; Coxhead & Byrd, 2007; Granger, 1998; Schmitt, 2004).

What corpus linguistics investigations have been able to capture through empirical investigations and further buttress with results showing statistical significance is a language phenomenon noted first by Jespersen (1924) and later by Firth (1957), who

observed that the words are typically associated with other words in a systematic manner. As a phenomenon it is characterized in Firth's words as, "You shall know the word by the company it keeps" (p.11). This property of a word to have its own "social network" prompted Boulinger (1976) to remark that learners produce not only phrases generated by their internal grammars but also phrases memorized as wholes. Nattingen and DeCarrico (1992) emphasized the importance of frequent word combinations that in writing, as well as speaking, may facilitate communication. The pervasiveness of the formulaic sequences in language use has led authors to propose different theories of language processing such as Sinclair (1991) or Hogue (2005), which remain subject to evaluation and testing. The frequency of multiword expressions supported empirically by the findings of corpus linguistics sparked an interest in the investigation of the topic multiword sequences, which is evidenced by four book length publications on the topic--Schmit (2004); Wray (2002); Meunier and Granger (2009); and Corrigan, Moravessik, and Wheatley (2009). In addition, Coxhead, Bunting, Byrd, and Morgan (forthcoming) are investigating the collocations and recurrent phrases of the words in AWL. Simpson-Vlach and Ellis (2010) compiled a list of most frequent formulaic phrases in academic prose. Martinez and Schmitt (2012) have created a phrasal expression list, a list of formulaic sequences suggested for L2 learners' receptive use.

### **Formulaic Sequences: Forms, Functions and Terminology**

The various forms formulaic sequences have in academic prose are reported in an extensive examination of English grammar conducted by Biber et al. (1999). The researchers offer a useful summary of multiword units entitled "Lexical expressions in

speech and writing” (pp. 987-1036). The summary briefly outlined below describes the versatility of forms of multiword sequences:

1. Sequences have semantic or structural function. These are word units like phrasal verbs (e.g., *put out* - extinguish) and prepositional verbs (e.g., *agree to*) that occur together and need to be regarded as a unit in respect to meaning and structure.
2. Sequences often referred to as idioms, which carry the meaning that is difficult if not impossible to derive from the individual meanings of the words that make up the unit (e.g., *birds of a feather flock together* - people of the same/similar character associate together).
3. Word combinations called collocations in which one vocabulary item prefers the company of another item rather than its synonym and each word retains its meaning (e. g., *obvious difference, obvious challenge*).
4. Multiword combinations that have lexico-grammatical associations. These are words that are strongly associated with a particular grammatical structure. For example, the verb *think* is strongly associated with *that*-clause; whereas the verb *want* is associated with a *to*-complement clause.
5. Word sequences that are “recurrent expressions, regardless of their idiomaticity and regardless of their structural status” (p. 990) are called *lexical bundles*, the term coined by the researchers. Lexical bundles are repeated in exactly the same form and sequence at a set frequency and over a range of different registers. Researchers consider lexical bundles of fewer than two words to be collocations and sequences of three or more words to be lexical bundles. Examples of lexical

bundles in academic writing include combinations of noun phrases followed by *of* (e.g., *the number of*) and other types of modifiers (e.g., *the fact that*); copular *be* followed by noun phrase (e.g., *is a matter of*) or adjective phrase (e.g., *is similar to*).

6. Sequences of semi-fixed structures with restrictions posited by grammar and vocabulary (e.g., modal + *well be*)
7. Sequences of verb + particle that combine freely but have a strong collocational association (e.g., *go down, come back*).
8. Multiword units consisting of two words of the same grammatical category coordinated by conjunctions *and* and *or* termed coordinated binominal phrases (e.g., *speech and writing, economic and political, positive and negative*).

In addition to having multiple forms, multiword combinations have multiple functions in academic prose. In academic prose they can serve as discourse devices to structure the text and connect meaning in a logical, temporal, and special order; to exemplify, qualify, or summarize (Nattingen & DeCarrioco, 1992).

Corpus linguistics research also reports on the contribution that formulaic sequences make to an academic text, in particular the contribution to its stylistic expressiveness. To illustrate, Gläser (1998) treats a range of sentence-length expressions, including proverbs, maxims, slogans, and quotations in various written genres and among them academic-scientific monographs (e.g., academic essays, research articles). The investigation has shown that the academic writing of expert writers<sup>2</sup> is marked by use of metaphors and allusions to proverbs and questions as well as modifications of formulaic

---

<sup>2</sup> Expert writers may include both L1 and L2 writers with expertise being defined as academic writing ability sufficient to publish in peer refereed, academic journals.



sequences. An example of this distinguishing feature of expert writing is found in the abstract written by Coxhead (2008) who announces the topic of discussion with the allusion to a famous quote by John Donne (1572-1631) by stating that “no word is an island” (p. 149). By so doing, she highlights the point that these words do not generally exist isolated from other words. She continues to exhibit the same writing style by stating further that “. . . words of a feather seem to flock together...” (p. 149) in this instance, modifying an every-day formulaic sequence (i.e., idiom) to point out that words tend to co-occur in texts.

Because of the variability in form and function of the multiword units, constructing a definition to account for variability within the phenomenon has been a challenging task (Schmitt, 2004; Wray, 2002). Due to the lack of a comprehensive definition, over 50 terms are applied in a description and study of multiword units (Wray, 2002). Amidst the negotiations of the naming system, the term *formulaic sequence* appears to be slowly establishing itself in the SLA literature through two veins: a) a line of research that relies on measures different from strict frequency measures and b) the handbooks on second language acquisition research, such as Ellis (2009), that utilize it as a cover term for this language phenomenon. The term is introduced by Wray (2002) and defined as follows:

a sequence, continuous, or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p. 9)

One of the strengths of Wray's (2002) definition is that it covers a wide range of sequences: a) sequences of various lengths and b) sequences found at both ends on the continuum of fixedness. This feature attributes to the applicability of the term to an examination of a range of multiword sequences in academic writing often found in studies that take a pedagogical perspective. It is because of these reasons the term formulaic sequences is appropriate for the study. However, it needs to be pointed out that terms, such as collocation, idiom, and lexical bundle, will be utilized in the body of the text as found necessary.

To summarize, the investigations from the corpus linguistics show that formulaic sequences are frequent and functional in academic prose and that they can contribute to self-expression and fluency in a text. The following section first highlights the benefits of using the formulaic sequences in ESL writing and then discusses the ways L2 learners deal with appropriate use in their writing.

### **Formulaic Sequences in L2 Writing: Benefits and Challenges**

The use of formulaic sequences performs several functions in academic prose, which can relate directly to the context of academic writing. Wray (2002) affirms that L2 learners' knowledge of formulaic sequences commonly used in a community, one of which is an academic community, will allow them to identify themselves with the members of this group, will ease their readers' reading effort, and will give them authority. Additional benefits to L2 learners' uses of formulaic sequences are noted by Coxhead and Byrd (2007):

- a) the [formulaic sequences] are often repeated and become a part of structural material used by advanced writers, making the students' task easier because they

work with ready-made sets of words rather than having to create each sentence word by word; (b) as a result of their frequent use, such [sequences] become defining markers of fluent writing and are important for the development of writing that fits the expectations of readers in academia; (c) these [sequences] often lie at the boundary between grammar and vocabulary; they are the lexicogrammatical underpinnings of a language that are so often revealed in corpus studies but much harder to see through analysis of individual texts or from a linguistic point of view that does not study language-in-use (pp. 134-135).

Both Wray (2002) and Coxhead and Byrd (2007) highlight the benefits of L2 learners' uses of formulaic sequences. Moreover, Wray (2002) warns that L2 adult learners' failures to see the vocabulary in terms of sets may negatively influence their literacy development, and Schmitt (2000) points out that our thinking about vocabulary in terms of single words needs to shift to thinking about vocabulary in terms of multiword combinations. The current research advances thinking in this direction.

## **The Usage and Acquisition of Formulaic Sequences by**

### **L1 and L2 Writers**

The previous section addressed briefly the naming issue in relation to the language phenomenon that certain words combine in a systematic manner with other words. Because of the variety of terms that probably refer to the same language feature, it is difficult to document fully what the research has shown about the usage of formulaic sequences in L2 writing. At this point, it is acknowledged that there may be other studies, in addition to those examined here that stress the issues involved in L2 writing.

### **The L2 Writers' Use of Formulaic Sequences**

The investigation of the L2 learners' text in terms of their use of formulaic sequences has taken a comparative analyses approach that has contrasted L2 advanced

student writers to a type of norm. For one line of research (De Cock, 2003; Erman 2009, Granger, 1998; Howarth, 1998; Paquot, 2008; Yorio, 1989), a native speaker represented the norm. The L2 students' texts were compared to the L1 students' writing (Altenberg & Granger, 2001; De Cock, 2003; Erman 2009, Yorio, 1989); corpora that combined L1 students' writing and L1 published writers' texts (Granger, 1998); or a corpus of L1 expert writers' texts both published and unpublished (Howarth, 1998). For another line of research (Hyland, 2008; Scott & Tribble, 2006), it was rather the expert writer (i.e., a writer whose work has been published or disseminated in some way) not the native speaker that represented this norm. The third stream of research (Neff, Ballesteros, Dafouz, Martínez, Rica, Diez, & Prieto, 2004; Neff van Aertselaer, 2008) considered both native speaker status as well as expert writer status in both L1 and L2 when attempting to describe the use of formulaic sequences in L2 writers' text.

The studies that report the use of an "expert writer" as a norm in the examination of the use of formulaic sequences of L2 writers' text, often fail to report the distribution of the published texts written by L1 and L2 expert writers. This failure makes it difficult to understand whether the "expert writer" norm represents L1 and L2 writers, L1 expert writers or, what seem unlikely, yet possible, L2 expert writers only. There have been examples of studies that compared L2 writers' compositions in terms of the frequency of use of formulaic sequences to the same corpus of texts (British National Corpus); yet, one study (Paquot, 2008) employed the term the native speakers' written production in the comparisons while the other (Scott & Tribble, 2006) referred to it as the expert writers' texts. However, despite the variation in terminology, this body of research was guided by the objectives of identifying the norms implicitly recognized by native speakers and

expert writers who may or may not be L2 writers and examining how L2 writing aligns with and departs from these norms.

The research results are in agreement on several features of L2 writing in terms of use of formulaic sequences. The research highlights the fact that ESL writers' use of formulaic sequences is problematic in multiple ways. First, ESL writers seem to fail to employ formulaic sequences where they are expected by logic. Second, when the ESL writers use the formulaic sequences, their writing may display a range of problems:

- A presence of errors of various types: lexical substitution (Howarth, 1998; Yorio, 1989), grammatical modification (Howarth, 1998; Yorio, 1989), pragmatic errors (Neff et al., 2004; Yorio, 1989) and semantic errors (Yorio, 1989).
- A manipulation of a limited number of formulaic sequences (Erman, 2009; Granger, 1998; Howarth, 1998; Neff et al., 2004; Scott & Tribble, 2006; Yorio, 1989).
- An overuse of a limited number of formulaic sequences (Granger, 1998; Paquot, 2008) that are often direct translations of the sequences in their native language (Altenberg & Granger, 2001; De Cock, 2003; Granger, 1998; Paquot, 2008) or by researchers assumed to be included in the L2 writing textbooks (Paquot, 2008).
- A production of non-target like formulaic sequences (Erman, 2009; Neff van Aertselaer, 2008) that may be a product of direct translations of common formulaic sequences appropriate for the academic prose in learners' native language (Neff et al., 2004).

- A utilization of formulaic sequences of reduced complexity (Neff, et al., 2004; Scott & Tribble, 2006) and those frequent in speech not writing (Scott & Tribble, 2006).
- A failure to use the formulaic sequences most often implicated in the texts of expert writers (Neff, et al., 2004; Neff van Aetselaer, 2008; Scott & Tribble, 2006), to use the ones that have a direction translation to English frequently enough (De Cock, 2003); and to use them in the correct location inside the sentence (Paquot, 2008).

While the research findings corroborate in multiple regards and furthermore build on one another, the results on the overall frequencies of formulaic sequences in L2 writers' text in comparison to the overall frequency in the expert writers' and native writers' text are contradictory. Howarth (1998) reports that L2 writers employ an alarmingly low number of formulaic sequences in their writing. He compared the use of formulaic sequences, in particular the collocations/idioms of the *verb + direct object* form of the expert native writers in academia to that of non-native masters' degree students. His examinations focused on phrases of two + words in a sequence. His results showed that advanced L2 writers employed astonishingly 50% fewer formulaic sequences than expert writers. This underuse of formulaic sequences in L2 writers' texts is corroborated by the results of Erman (2009), Granger (1998), and Neff van Aertselaer (2008).

On the other hand, Scott and Tribble (2006) and Hyland (2008) report that L2 learners' frequency of use of formulaic sequences does not lag behind the expert writers. On the contrary, in case of the subjects in Hyland's study, they exceed the number of

frequencies in the academic texts of expert writers. A closer look at the two studies and a subsequent comparison of Howarth (1998) to Scott and Tribble (2006) and Hyland (2008) reveal the cause of results that conflict.

Scott and Tribble (2006) attempt to identify the extent to which the work of apprentice writer, masters' theses written by advanced Polish speakers majoring in English literature, is similar to and different from that of expert writers found in the British National Corpus (BNC), a 100 million balanced corpus of written and spoken English, in terms of their use of formulaic sequences, in particular lexical bundles of three- and four-word combinations. The researchers make two comparisons--one with the entire BNC academic corpus and another with a selected few articles from literary journals found in the BNC corpus. The comparison showed that L2 writers' employment of formulaic sequences in their writing did not lag behind that of expert writers in terms of the overall frequency.

The comparison revealed that as the complexity of the lexical bundle increases, as measured by the number of words in the bundle, the overlap between the most frequently used lexical bundles decreases. Additionally, the strings that are typical for published authors are less common in L2 writers, and L2 writers employed formulaic sequences found in the usage more commonly associated with speaking rather than writing (e.g., *a kind of*). More importantly, it is not that the writers used fewer formulaic sequences, but that they failed to utilize those formulaic sequences most often implicated in evaluations offered by expert writers (Biber, 1988), such as *anticipatory-it* as in *It is hard/possible/true*.

Hyland (2008) set out to explore the forms, structures, and functions of formulaic sequences in three corpora: expert writers' research articles, L2 doctoral theses, and L2 masters theses. The target of investigation was four-word formulaic sequences. The results of the comparison in frequency between the corpora under investigation showed that L2 learners in particular L2 masters theses contained approximately twice as many formulaic sequences than the expert writing. Similarly, doctoral dissertations written by L2 learners utilized more than the expert text.

Scrutiny of the methodologies and predetermined requirements for the formulaic sequences investigated in the study reveals that the reason for contrasting results lies in the different approaches taken to investigation of overall frequency of formulaic sequences in L2 writing in theses and dissertations and expert writing. Howarth (1998) examined all two + combinations of a verb + object sequences that allowed for inclusion of a possibly higher number of word combinations to be constructed, which allowed for a more comprehensive examination of the combinations. Scott and Tribble (2006) and Hyland (2008) based their examination on the sequences of a limited number of words in a sequence thus, excluding all the two-word sequences as well as those five and higher. Based on the results provided by the aforementioned studies, it can be concluded that L2 learners rarely employ verb + object collocations but rely on three and four-word combinations that constitute lexical bundles more frequently than the expert writers. This is particularly true if frequency, and not the accuracy and appropriateness of use, is the subject of investigation.

To summarize, the research that attempts to describe the use of formulaic sequences by L2 writers through a means of comparison with native and expert users of



formulaic sequences illuminates the issues L2 learners experience with the employment of formulaic sequences that relate to their background as L2 writers. In addition to struggling with issues related to their second language background, research suggests that L2 learners wrestle with the requirements imposed on them by the problems induced by their lack of experience in writing in the target language, problems that are similar to those of their native English-speaking peers. The studies conducted by Neff van Aertselaer (2008) and Neff et al. (2004) have been instrumental in this regard. They compared native expert, native novice, and L2 writers' texts, which allowed for multiple comparisons between the groups. The comparison between the writing of native novice and L2 novice writers showed similarities in manner of expression, which are different from those of expert writers'. To be more specific, Neff et al. (2004) investigated the employment of formulaic sequences used as subjective and objective stance markers in argumentative essays of the three groups. In terms of frequency, L2 writers' use of subjective stance markers was similar to that of the novice L1 writers. That is, both groups frequently employed construction of first person singular and the verb of mental process (i.e., *think*) or perception (i.e., *feel*) in contrast to native writers who used fewer of such expressions. Significant differences were found in the use of objective writer stance by using various formulae involving *it is + adjective* and the agentless passive patterns. The L2 writers overused agentless passive patterns and also patterns denoting obviousness (*it is clear that*) instead of modalized constructions (e.g., *It may be true that X, but Y*) of expert writers. The researchers conclude that certain constructions in particular those of higher complexity that combine both hedges (e.g., modal *may*) and boosters (e.g., *but*) represent a challenge for novice and L2 writers alike.

Similarities in L1 novice writing and L2 novice writing were emphasized by Neff van Aertselaer (2008). She conducted a corpus-based contrastive analysis study of the usage of sequences (e.g., expression of certainty, attitude, and impersonal presentation) by EFL students who are native Spanish speakers in comparison to those used by novice English writers, expert English writers, and expert Spanish writers. The analysis revealed that EFL writers, similar to their novice L1 counterparts used forceful adjectival phrases (e.g., *It is obvious*) and adverbs (e.g., *obviously, clearly*), some not found or found with very low frequency in the text of the expert English writers. The researcher remarks that both L1 and L2 novice writers display an incomplete mastery of formulaic sequence use for the purposes of argumentative essay writing.

To summarize, the research that attempts to describe L2 writers' practice in formulaic sequence use has shown that L2 learners' use of formulaic sequences in writing diverges in multiple ways from expert writer norms but at the same time exhibits some similarities with those of the novice L1 writers. Not only do L2 writers fail to employ formulaic sequences within the text in places where they are expected by logic, but when they actually attempt to use the phrases, they also fail in several ways. For example, the forms are deviant or non-native and are representative of a spoken rather than a written register. In addition, when they are of the correct form and in the correct register, they can appear in unexpected locations within a sentence. Based on findings reported by the studies previously examined, it seems that L2 learners, in order to use formulaic sequences in an expert-like manner to become a part of the academic community, have two large hurdles to overcome: one is created by the second language background and the

other by a lack of experience in L2 writing, the latter being similar in many ways to that of L1 novice writers.

In order to illustrate the intricacies involved in the usage of formulaic sequences in general and to highlight the difficulties with which L2 learners are faced, the discussion very briefly turns to an examination of use of formulaic sequences by L1 writers.

### **L1 Learners' Use of Formulaic Sequences**

The research has pointed out that novice and expert English L1 writers differ in the frequency with which they use the target formulaic sequences. To illustrate, Neff van Aertselaer (2008), in an examination of L1 and L2 novice and expert writing of argumentative essays, showed that the L1 novice writers utilized frequently forceful adjectival phrases and adverbs as well as doubtful passive constructions that showed a lack of mastery in the construction of balanced arguments that force a reader, rather than convince a reader, to agree with their position.

The Neff et al. (2004) study, in another comparative text analysis that allowed for a comparison of L1 novice and expert writers, has shown that native writers while using the same formulaic sequences use them to achieve various purposes in the text. Neff et al. examined the stance markers in L1 and L2 novice and expert writing and found that the same formulaic sequence used by L1 novice writers as a metadiscourse marker to signal a shift in topic is used by the expert writers to show agreement with the source used as reference in argument construction.

Additional examples of how L1 writers' depart from the expert writers in the use of formulaic sequences, and lexical bundles specifically, comes from Cortes (2002, 2004) who examined the L1 and expert writers use of formulaic sequences in academic writing. Cortes (2002) models her study after Biber et al. (1999), thus analyzing a corpus of freshman composition writing with a goal to find out the most common four-word lexical bundles produced by freshmen university students and to establish a comparison between the four-word lexical bundles most frequently used in academic prose and conversational lexical bundles originally identified by Biber and Conrad (1999). The researcher first conducted a grammatical and functional analysis of the target formulaic sequences in the corpus of novice L1 writers and an analysis of the functions performed by the expert writers. The initial investigation of the corpus of freshmen composition writing showed a structural similarity in lexical bundles between novice and expert writers. However, a closer examination highlighted the fact that despite the similarity in the grammatical structures of lexical bundles, there were differences in the function lexical bundles performed in the expert and novice writing. The novice writing was characterized by a frequent use of lexical bundles denoting location and time frame.

In another study, Cortes (2004) compared the frequency and function of lexical bundles in the written output of published authors and L1 students' writing in history and biology. The lexical bundles were extremely frequent in the corpora of published articles from popular American journals in the disciplines; the use of lexical bundles by the students was rare. The students were all native speakers of English at different university levels. She concluded that the exposure to the lexical bundles did not readily transfer to the students' active production of lexical bundles in writing. Furthermore, on the rare

occasions that the students did use the lexical bundles, their use differed from the use of the published writers. The researcher attributed the lack of the use of sequences to the lack of formal instruction that students in the disciplines have on the frequency and function of the expressions.

While the research points out the benefits of L2 (and L1 writers') use of formulaic sequences and the challenges all learners have in using the sequences, more needs to be known on the processes involved in learning the formulaic sequences for the purposes of production in writing; that is, more needs to be known about the roles of incidental and explicit learning of formulaic sequences (Li & Schmitt, 2008). The research conducted in the formulaic sequence acquisition for purposes of writing in L1 still needs to explain how native speakers transcend to the expert-like use of formulaic sequences (Cortes, 2006; Schmitt & Carter, 2004). Cortes (2006) represents an attempt to fill the gap in L1 studies of the effects of explicit teaching of formulaic sequences on L1 writers.

Cortes (2006) focused on direct instruction of a limited number of selected categories of formulaic sequences, that is, lexical bundles in a writing-intensive history class and on the analysis of simple expressions favored by the students instead of lexical bundles. The study investigated the effects of teaching four-word lexical bundles on the frequency of use of these bundles in the learners' writing by means of a pre-/posttest and students' perceptions of the relevance of the bundles in their academic writing. The study results showed no major improvement in the frequency and the variety of the use of lexical bundles between the pre- and posttreatment written assignments. The examination revealed a rare and uneven use of the target formulaic sequences with a few of the new bundles used prior to instruction and some lexical bundles not used by the students after

the instruction. A survey of students' views of the use of lexical bundles revealed that their awareness about the use, frequency, and function of lexical bundles in published writing increased; however, the analysis of the students' papers revealed that the increased level of awareness also transferred over to students' actual use of the bundles in their writing. The analysis of students' papers in search of alternative words or expressions used to convey the functions of lexical bundles revealed that students favored simple expressions.

### **L2 Learners' Acquisition of Formulaic Sequences**

One of the objectives of the study of formulaic sequences in L2 writers is an investigation of how formulaic sequences are acquired. The following section discusses some studies that have examined the L2 writers' development in terms of use of formulaic sequences.

Li and Schmitt (2008) is an example of a longitudinal case study conducted to observe how the use of formulaic sequences in the writing of a Chinese graduate student developed over a period of one academic year. The researchers examined multiple pieces of the participant's written work (i.e., eight essays and a dissertation) and conducted interviews to collect information on the learner's previous experiences with English and strategies used in writing in English. The researchers found out that the subject learned a large number of new formulaic sequences. To be specific, she learned 166 formulaic sequences, which, added to 153 sequences she already knew, accounted for fifty percent of the total number of word combinations she had learned. The analysis of the papers showed the learner's tendency to employ a limited number of formulaic sequences, a

finding that corroborates those of Granger (1998), Horwarth (1998), and Yorio (1989). However, the most notable finding of their study is probably that the subject's learning of formulaic sequences was taking place incrementally, which is similar to the acquisition of single words (e.g., Nation, 2001). The improvements the learner made over an academic year were in the area of appropriateness of the use of formulaic sequences. The increase in control over formulaic sequences was mainly of partially known phrases becoming more appropriately used, suggesting that learning formulaic sequences was not necessary but polishing the knowledge of the sequences already familiar to the learner to a certain degree was.

Given that L1 research on formulaic sequence acquisition for the purposes of writing is in its early years, the field has turned to the general field of vocabulary acquisition for guidance on how to teach L2 learners' formulaic sequences for production (Jones & Haywood, 2004; Schmitt, Dorneyei, Adolphs, & Durow, 2004). Given the evidence that suggests formulaic sequences are acquired incrementally (Li & Schmitt, 2009; Schmitt et al., 2004) similar to the way vocabulary is acquired (e.g., Nation, 2001), this may be a useful approach to investigation of the learning of formulaic sequences.

In the area of vocabulary instruction, two main approaches have been suggested: indirect and explicit. Indirect vocabulary teaching allows for incidental vocabulary learning which refers to learning the words through sheer exposure to the target language "when one's attention is focused on the use of language, rather than the learning itself" (Schmitt, 2000, p. 116). Explicit vocabulary instruction allows for explicit learning that takes place when focused attention is paid to learning target vocabulary. The literature on vocabulary teaching and learning (e.g., Coady, 1997; Folse, 2008; Hulstijn, 2001;

Nation, 2001; 2005; Schmitt, 2000) regards both approaches important in vocabulary instruction, but recommends that explicit teaching methodology be employed for the words that are deemed important and the time for learning of these words is limited.

However, it would be premature to conclude that every explicit teaching vocabulary intervention has similar effects. In fact, explicit instruction that uses decontextualized exercises may be less effective than an activity which promotes implicit learning (Elly & Mangubhai, 1983; Krashen, 1989). The explicit teaching that can be effective is provided in a systematic manner and in meaningful contexts (Laufer & Shueli, 1997; Paribaht & Wesche, 1997; Zimmerman, 1997) to allow for multiple encounters with the target word (Laufer & Shueli, 1997; Paribaht & Wesche, 1997; Zimmerman, 1997). It also employs a variety of instructional techniques (Paribaht & Wesche, 1997; Zimmerman, 1997) including deeper level processing (Laufer & Shueli, 1997; Paribaht & Wesche, 1997).

An advantage of explicit instruction is that it “focuse[s] attention on forms and meanings in the input” (Schmidt, 2001, p.10) and, thus, can facilitate *noticing*, the first of the three conditions, that are according to Nation (2001) necessary for a word, and by extension a formulaic sequence, to be learned. The second condition is *retrieval*, which can be receptive (i.e., perceiving the word form and retrieving its meaning when the word is seen in written text or heard in a conversation) or productive (i.e., having a desire or need to communicate the meaning of a word and having to recall the spoken or written form for speaking or writing purposes). Once a word has been noticed and retrieved, the third major process that encourages learning is *generation*. It occurs when a previously encountered word is seen, heard, or used in a slightly different environment.



In order for a learner to know a word, and, by extension, a formulaic sequence, according to Nation (2001, 2005), he or she needs to know various aspects of *knowing* the word, which fall into one of the following three groups: form, meaning, and usage. The aspects within the groups require both productive and receptive knowledge. The former is often referred to as passive knowledge, and is the kind of knowledge a learner needs in the acts of listening and reading. The latter, however, is also called active knowledge, which a learner needs in order to use a word in speaking and writing.

Under the assumption that the processes involved in learning a formulaic sequence are similar to the processes involved in learning a word and that the direct teaching of vocabulary may expedite the learning of formulaic sequences, few studies (Jones & Haywood, 2004; Schmitt, et al., 2004) have been conducted to examine the facilitative effects of explicit instruction of formulaic sequences on ESL learners' ability to produce them in written tasks.

Schmitt et al. (2004) represents an attempt to measure the receptive and productive knowledge of 20 selected, academically based formulaic sequences in EAP students of fairly high proficiency studying to enter postgraduate studies at British universities. The instruction extended over a period of 2 months for 62 and three months for 32 students. In a semicontrolled environment the amount of exposure or instruction of the target formulaic sequences the participants received was not controlled (the instructors were asked to draw learners' attention to each formulaic sequence at least once within a semester). The students' improvement in knowledge of formulaic sequences was measured by means of a pre- and post- receptive and productive knowledge test. The study results showed that the participants came to the program with

a substantial knowledge of formulaic sequences. This knowledge improved over the course of the instruction on formulaic sequences used productively and receptively; however, given that the study employed no control group, it is not clear whether the intensive program designed as such facilitated the learning of formulaic sequences or whether the explicit instruction, consisting of at least one exposure to the target formulaic sequences, was sufficient for learning the sequence (in terms of receptive or productive skills). Even though the study gives some persuasive evidence of at least some degree of increased productive knowledge of formulaic sequences, as demonstrated by the ability to complete a cloze item successfully, it does not provide evidence of the ability to use the formulaic sequence freely in written communication.

In an exploratory study, Jones and Haywood (2004) investigate the effects of explicit instruction on formulaic language on L2 writers in an EFL context during a 10-week EAP pre-university course. The study was conducted to meet three conditions necessary for vocabulary learning (i.e., noticing, retrieval, and generation) suggested by Nation (2001). The instruction was provided in a reading and writing class. In the reading class, the formulaic sequences were explained as a phenomenon and the course instructors used awareness raising techniques to help learners notice the sequence in the reading materials. The reading class also offered opportunities to practice the formulaic sequences in controlled environments.

In the writing class, the students reviewed the awareness raising activities and were “encouraged to try” (Jones & Haywood, 2004, p. 275) by including them in their own compositions. The activities integrated in the instruction offered in a writing class asked students to examine the pragmatic functions of the target formulaic sequences in

different contexts, to classify them based on their lexico-grammatical patterns, to supply a correct formulaic sequence from a list in an appropriate location in a text, and to produce concordance texts by using a concordancing program. These types of in-class activities, which appear to encourage learners to manipulate readily available formulaic sequences and to focus on an analysis of their forms and functions, seem to encourage the development of receptive knowledge of formulaic sequences. The study results show learners' marked gains in noticing, modest gains in production in a controlled situation (i.e., C-test), and no gains in production in an uncontrolled situation (i.e., essays). Based on the learners' modest gains in production in a controlled situation, the researchers conclude that some learning of the formulaic sequences has occurred as a result of the explicit teaching approach.

The Jones and Haywood (2004) study findings seem problematic for several other reasons: First, the groups were taught by two different instructors who could have employed two different teaching styles. Second, only 2 weeks was allowed for explicit instruction on formulaic sequences prior to the collection of data on the composition used as a post-test. Third, the small number of data coming from a relatively small sample was depleted further with several students not completing the posttest, which did not allow researchers to reach any conclusive results. Fourth, little attempt was made to control for the task effect on the two measures of students' production skills: the pre-/post- C-tests differed in length and in the sequences they targeted and the pre-/posttest essays required different organizational patterns.

There is additional evidence that explicit teaching can facilitate the learning of formulaic sequences. First, Li and Schmitt (2009) report that for their participant, explicit

instruction was one of the major sources of acquisition for new formulaic sequences. Mainly, according to the participant's account, 31% of the total number of different types of newly acquired formulaic sequences used in writing came from the explicit instruction. In addition, Yorio's (1989) investigation revealed that the learners showed accurate use of those formulaic sequences that were explicitly taught in their writing classes.

The literature review has so far attempted to point out that L2 learners may need direct instruction in order to attain the levels expected in academic writing of a mature and advanced users of formulaic sequences. To add to the previous studies calling for teacher assistance in the development of L2 formulaic knowledge for the purposes of L2 academic writing, the paper briefly turns to an examination of a recent study by Ellis, Simpson-Vlach, and Maynard (2008) and the most recent attempts at identifying useful formulaic sequences (Martinez & Schmitt, 2012; Simpson-Vlach & Ellis, 2010).

Ellis et al. (2008) examined how different aspects of lexical patterning affect the accuracy and fluency in the processing of these formulaic sequences in native speakers and advanced ESL learners. Three factors of lexical patterning were investigated: length, frequency, and *MI* (i.e., a statistical measure of how much the words in a collocation cohere or are found in collocation) of formulaic sequences. The sequences were initially identified as sequences with educational and psychological validity. The study results show that both native speakers and non-native speakers are sensitive to the formulaic sequences; however, they are sensitive to different aspects of *formulaicity* (i.e., relative fixity between words in a formulaic sequence); whereas, native speakers' processing of word sequences is affected by the *MI*, advanced L2 learners' language processing is affected by the formulaic expression frequency. It seems that native speakers do not only

differ in the way they use formulaic sequences but also in the ways both groups process them.

Based on the study findings, Ellis et al. (2008) suggest that L2 learners need support in learning the formulaic sequences. They are concerned that their participants, advanced learners of English with 10 years of experience learning the language and some in immersion programs, still processed the formulas in a non-native like manner. That is why, they conclude, that L2 learners need direction and guidance in recognizing and becoming attuned to formulaicity in order to start processing them in a native-like fashion. They go further to suggest that assistance needs to be provided in the language classes, specifically English for Academic Purposes (EAP) classes. To assist L2 instructors in the selection of formulaic sequences to teach, the researchers identified through a combination of quantitative and qualitative criteria, corpus statistics, linguistic analysis, psycholinguistic processing metrics, and instructor insights, an Academic Formulas List (AFL).

An Academic Formulas List, developed by Simpson-Vlach and Ellis (2010) is an empirically derived list produced after an examination of a 2.1 million words of written and 2.1 million words of spoken academic discourse across various disciplines (humanities and arts, social sciences, natural sciences and medicine, and technology and engineering). The researches applied *frequency*, a statistical measure of cohesiveness, in combination with validation and prioritization studies, to determine which formulaic sequences are worth teaching to create a list of formulaic sequences of three to five n-grams suggested for instruction. The list consists of a Core formulas list, 207 formulaic sequences that are shared between academic written and academic spoken language and

two separate formula lists that separately present the 200 formulaic sequences frequent in written and 200 formulas frequent in spoken form. The formulas are classified by pragmatic-linguistic function to facilitate the instruction in the EAP curricula and ranked according to the empirically derived psychologically valid measure of utility, called “formula teaching worth” (FTW). The list compilation procedures, as well as the number of sequences included in the list, suggest high pedagogical utility.

The need for support in learning and teaching of formulaic sequences is evidenced in another attempt at creation of a pedagogically useful list of formulas. Martinez and Schmitt (2012) combined quantitative and qualitative analytical procedures to extract 505 formulaic sequences of two- to four-word combinations with high frequency from the BNC corpus. The researchers first extracted two, four-word formulaic sequences, occurring at least 787 times in the corpus under investigation. The primary investigator, guided by the selected criteria proposed by Wray (2008), identified and ordered 505 formulaic sequences for pedagogical and assessment purposes. The list, as the researchers state, has been created with an intention to assist learners in the receptive acquisition of formulaic sequences. Because the list is aimed at building a receptive knowledge of formulaic sequences and the present study attempts to measure the gains in the productive, as well as receptive, knowledge of formulaic sequences, the AFL developed by Simpson-Vlach and Ellis (2010) is employed.

To summarize, the evidence from corpus linguistics shows that multiword sequences are frequent and functional in academic prose. This evidence suggests that in order for L2 learners to “sound” academic in their writing, gaining control over the use of formulaic sequences is important. The research on explicit instruction of formulaic

sequences, though limited, provides some support for explicit instruction in learning of the formulaic sequences. The evidence from psycholinguistics suggests that non-native speakers not only differ in their use, but also the processing of the formulaic language. The research also stresses that teacher intervention via explicit instruction is needed. Moreover, by creating a list of AFL specifically focusing on academic prose, the research on formulaic language processing suggests that formulaic sequences need to be treated directly in L2 writing classes for university-bound students. It is evident that more research is needed in order to make claims that explicit instruction can indeed assist L2 speakers in learning new formulaic sequences and improving the mastery over the sequences that are partially known.

The present study builds on Jones and Haywood (2004), but it attempts to correct the limitations of the previous study to better understand the effects of explicit instruction on the L2 writers' abilities to produce formulaic sequences. It is designed to (1) employ one instructor to teach both groups of participants; (2) devote one entire term to explicit instruction of formulaic sequences used in one type of writing, argumentative essay writing; (3) use a larger study population; (4) allow students who miss class to make up the missed assignments; and (5) control for the task effects on the pre-/post-test measures on production tasks.

The design of the present study departs from Jones and Haywood (2004) in three major ways: (1) explicit instruction of the formulaic sequences is integrated in the writing class instruction only as opposed to distributing the segments of instruction between a reading and a writing class; (2) the effects of explicit instruction of the formulaic sequences on the students' ability to identify them in a text are not examined; and (3) the

target formulaic sequences are different from those selected by Jones and Haywood who referred to Biber et al. (1999) as a source. This study attempts to capitalize on the most current findings of Simpson-Vlach and Ellis (2010) to select from the Academic Formulae List, the list of frequent and pedagogically useful list of phrases. In addition, the present study also takes into account the information provided by the literature of content-based instruction (e.g., Christison, forthcoming; Murray & Christison, 2010) and collocational research (i.e., Erman, 2009). Both lines of research emphasize that learners can discuss effectively a topic when appropriate vocabulary on the topic is utilized. Specifically, the literature of content-based instruction, which currently views vocabulary in terms of single words, utilizes the expression *content-obligatory* vocabulary and defines it as the vocabulary which is necessitated by the topic of discussion. For example, an L2 learner asked to write a paper on the topic of different forms of government can discuss the topic effectively only if the words such as *government, democracy, autocracy, monarchy* are utilized in the text.

In the area of research on formulaic language, in particular collocations, Erman (2009) offers a comprehensive discussion of the identification and definitions of collocations from different theoretical frameworks to propose an alternative view of collocations and argue that certain word combinations are motivated by a specific topic. To illustrate, if an L2 learner is writing a paper on environmental issues he/she is expected to use the word combinations such as *global problem, environmental awareness, protect the environment* in order to construct mature and expert-like prose. The researcher refers to them as topic-induced and points out that they are necessary and expected in well-written prose on a specific topic.



A parallel can be drawn between the learners discussing topics in a content area to the learners discussing topics in an ESL writing classroom. Both groups of students are asked to write on a specific topic, with the difference being that one group of students does it in a content area and the other in the context of ESL writing, and ESL writers also need content vocabulary that the specific topic of writing obligates them to employ. They need to use the topic-induced vocabulary to sound more mature and advanced in their writing, which is a quality that is highly valued in writing (Grobe, 1981; Nielsen & Piche, 1981).

## CHAPTER 3

### METHODOLOGY

The present chapter offers an overview of the research design, study procedures, including the creation of materials for explicit instruction, and the design of instruments used in the study focused on revealing the possible effects of explicit instruction of formulaic sequences on L2 writers. The questions that motivate quantitative analyses are the following:

1. Is there a significant difference in students' abilities to *produce*, in a controlled situation (i.e., C-test), formulaic sequences frequent in academic prose (i.e., selected AFL [Simpson-Vlach & Ellis, 2010]) between the students who receive explicit instruction and those who do not?
2. Is there a significant difference in students' abilities to *produce*, in a controlled situation (i.e., C-test), formulaic sequences in topic discussion between the students who receive explicit instruction and those who do not?
3. Is there a significant difference in students' abilities to *produce*, in an uncontrolled situation (i.e., an essay), formulaic sequences frequent in academic prose (i.e., selected AFL [Simpson-Vlach & Ellis, 2010]) between the students who receive explicit instruction and those who do not?
4. Is there a significant difference in students' abilities to *produce*, in an

uncontrolled situation (i.e., an essay), formulaic sequences in topic discussion between the students who receive explicit instruction and those who do not?

5. Is there a significant difference in the overall quality of students' essays as measured by the trained essay raters' judgments between the students who receive explicit instruction and those who do not?

In addition, the study attempts to conduct an in-depth inquiry into the strategies applied by students who used formulaic sequences in topic discussion and/or academic formulaic sequences more frequently and with greater accuracy than their peers receiving the same type of instruction. This is achieved through post-treatment interviews with a selected group of students from the treatment group. The students are asked two types of questions: One category of questions focuses on identifying the strategies learners apply in identification and production of the target formulaic sequences while completing specific tasks. These questions are created for each learner individually based on his/her performance on these tasks. The other category of questions explores the strategies learners use in the study and production of the target formulaic sequences for the purposes of writing.

### **Overview of the Research Design**

The present research uses a quasi-experimental design in which the study participants are assigned to experimental and control groups based on the class in which they are enrolled. The study was conducted in writing classrooms in a university intensive English program (IEP) that focuses on writing argumentative essays. It is the first course in a series of two courses that attempts to prepare students for study at the

university level and helps them develop the skills required for academic writing.

Instructional periods in the IEP are divided into terms of eight weeks, with two terms occurring each semester.

The treatment group received explicit instruction on formulaic sequences, instruction that was aimed at raising L2 writers' awareness of formulaic sequences, as well as the use of formulaic sequences in their own writing; the participants were quite possibly also exposed to the specific target formulaic sequences in questions through reading and in class discussion. The control group received no explicit instruction on formulaic sequences but focused instead on other writing-oriented activities. The group was exposed to the target formulaic sequences only through reading and possibly in class discussions. The control and experimental groups were given the same composition assignments.

Other features of research design include the fact that both groups received the same amount of instruction time for one term or eight weeks. Both groups also followed the same syllabus, read and discussed the same reference materials, and were taught by the same instructor to reduce the effects of the teacher variable on the results.

Besides gathering data for quantitative analyses, the design of the current study allows data to be collected for qualitative analysis because qualitative data will "provide depth and detail" (Patton, 1990, p.22) to the investigation of the effects of explicit instruction of formulaic sequences on L2 writers. The qualitative data were collected through individual interviews with a selected group of learners from the treatment group based on their performances on the tasks used for quantitative data elicitation. The interviews were semi-structured and approximately 15 minutes long. The interviews

allowed for an examination of the learners' viewpoints (i.e., emic perspective) on the outcomes of explicit instruction of formulaic sequences in L2 writing classes and also more detailed information about the strategies learners applied to learning how to use formulaic sequences in their writing.

### **Selection of Target Formulaic Sequences**

The formulaic sequences utilized in the study constitute a subset of formulaic sequences from the Academic Formulas List (Simpson-Vlach & Ellis, 2010) as well as the formulaic sequences used in the discussions of topics examined in the writing course. The formulaic sequences were selected based on two main criteria. First and foremost, the formulaic sequences needed to be useful to students and worthwhile to learn. Second, they needed to be present in the course environment, so they needed to be a part of academic discourse and relevant to discussions of topics examined in class. Based on the criteria of relevance to the class instruction and utility, the formulaic sequences selected for explicit instruction were those identified in the readings students complete in preparation for writing, which included background readings on the selected topics and essay models in the target genre. A large portion of the material had been preselected by the curriculum developer at the IEP and the remainder of readings by the course instructor with some minor input from the researcher.

The formulaic sequences offered in Core and Written AFL (Simpson-Vlach & Ellis, 201) lists have been empirically derived and identified as frequent and functional in academic writing--in other words, the type of writing university and university-bound L2 students are expected to emulate (For more details on the features of academic writing

refer to Chapter 2, pp.31-32.). They have been organized into sections by the pragmatic functions they perform in academic writing (e.g., exemplification, stance expression, compare/contrast, etc).

The process of identification of the formulaic sequences listed in AFL (Simpson-Vlach & Ellis, 2010) within the selected texts currently represents a time-consuming task because, to the researcher's knowledge, an on-line and free-of-charge software tool for locating AFL formulaic sequences is not offered at this time. The unavailability of such software is why the process of identification in the current study is represented a multistep procedure that requires the use of at least two computer software programs that are on-line and free-of-charge. The procedure was conducted as follows: First, the course readings were scanned into the text files, proofread, and saved on the computer. Second, the list of Core AFL formulaic sequences and the list of Written AFL formulaic sequences were transferred from Appendices A and B. They were located in the Supplementary Data file of the on-line version of the academic journal and were saved on the computer. Third, the readings and the AFL sublists were submitted to the two lexical analyses via two different computer software programs: a) Text-Lex Compare v.2. 2 (Cobb, 2010); and b) Microsoft Windows version 2007.

Text-Lex Compare v.2.2. (Cobb, 2010) is a corpus linguistics software tool that allows comparison of lexis, both at the word and phrase levels, of two or more texts, lists, and texts and lists. The output it generates at the phrase level offers phrases no larger than three words (i.e., three n-grams). This program was found to be a suitable software tool for the initial investigation for the present study because it was designed to compare phrases in texts to the phrases in lists to identify those that are common to both.

Subsequent analysis needed to be conducted to identify the formulaic sequences larger than three-word sequences.

To obtain the information regarding which of the three-word formulaic sequences from the Core and which of the three-word formulaic sequences from the Written AFL formulaic sequences are present in the readings for the course, the texts and the two complete sub-lists of AFL formulaic sequences were individually imported into in the Text-Lex compare v.v.2. The program allows two methods of data input: Method 1 requires that the texts and lists be copied and pasted into appropriate locations, and Method 2 asks for files to be loaded from the computer database to the program. Either method can be used for the present analysis; however, because Method 1 appeared to be a more time-efficient approach than Method 2 for the format of the data of the present study, it was selected. The study procedure was conducted as follows: the readings were entered in the window entitled “Old Text,” and each AFL list is entered separately as data in the window entitled “New Text.” The Text-Lex Compare program generated four lists of phrases: “Unique to old,” a list of phrases that appear in the “Old” text or the class readings; “Unique to new,” a list of phrases exclusive to the “new”, that is, the Core or Written AFL arranged by frequency and alpha listing; and “Shared,” a list of phrases that appear in both the readings and the AFL list. It is the information generated by the program in the list entitled “Shared’ that was used for further analysis for the purposes of the present study.

The list of formulaic sequences that are shared between the readings and the Core AFL and the list of formulaic sequences that are shared between the readings and the Written AFL were imported into the EXCEL program and aligned with the original

complete Core and Written AFL sub-lists imported from the supplementary materials folder of the Simpson-Vlach and Ellis (2010) on-line publication version. Through the process of alignment, all three-word formulaic sequences of the AFL sub-lists that were present in the readings were identified. The procedure of aligning the formulaic sequences revealed that because Text Lex Compare v.2.2. (Cobb, 2010) had not been developed to compare phrases more complex than those of three words and the phrases in the Core AFL and Written AFL range from three (e.g., *in terms of*) to five-word phrases (e.g., *from the point of view*), some formulaic sequences generated in the list of “Shared” formulaic sequences represented incomplete realizations of the target form. For example, one of the phrases in the list of shared formulaic sequences between the readings and the Written AFL is “*a result of*”; whereas, the target Written AFL formulaic sequences is “*as a result of the.*”

In order to identify the Core and Written AFL formulaic sequences more complex than three-word combinations, an additional lexical analysis needed to be performed, as noted previously. The subsequent analysis was conducted using the Microsoft Word version 2007 program with its search feature “Find” and performed as follows: The course readings were imported in one Microsoft Word version 2007 document entitled “Complete readings.” The search was conducted by activating the “Find” feature of the program and by entering the three plus Core and Written AFL formulaic sequences as search criteria one at a time. The program is designed to highlight the target formulaic sequence within the text under examination. Each time the target formulaic sequence was located in the text, the researcher examined the context to determine whether the pragmatic function of the formulaic sequences in the AFL lists matched the pragmatic



function of the sequence highlighted in the text. If the pragmatic functions matched, the formulaic sequences were added to the lists of the Core and Written AFL formulaic sequences located in the readings.

The Microsoft Word version 2007 program with its search feature “Find” was used for an additional investigation of pragmatic functions of the three-word formulaic sequences. The formulaic sequences identified in the texts that perform the same pragmatic function as those of the Core and Written AFL formulaic sequences were added to the lists of the Core and Written AFL formulaic sequences located in the readings. Through this examination several three-word phrases identified by the Text Lex Compare (Cobb, 2010) program had to be eliminated. For example, the Text Lex Compare identified, in the readings under investigation, a phrase “*appears to be*” which also occurs in the same form (i.e., “*appears to be*”) in the Written AFL list (Simpson-Vlach & Ellis, 2010) in the group of Stance Markers and is narrowly defined as a Hedge (e.g., The researcher *appears to be* avoiding to . . . ). Upon further investigation of the pragmatic function of the phrase “*appears to be*” by using the Microsoft Windows version 2007 tool “Find,” the phrase was found in the reading materials where it is clear that the function it performed was far from a hedge as can be seen in the following example found in the texts under investigation: “*These word groups may contain a noun that appears to be the subject but in fact modifies the subject.*”

The Microsoft Word version 2007 program with its search feature “Find,” was also useful in another way. As it searches the readings for the potential occurrences of the Core and Written AFL formulaic sequences of three plus words, it allows the researcher to verify whether the three-word phrases generated by Text Lex Compare v.v.

2 (Cobb, 2010) represent complete forms as they appear in the AFL sub-lists (Simpson-Vlach & Ellis, 2010) or as they are embedded in larger AFL phrases constructed of four and five words. For example, the phrases “*the form of*” and “*in the form*” are embedded in the target formulaic sequences realized as “*in the form of.*”

Additional benefit of using the Microsoft Windows “Find” tool is that it could point to the exact location of the formulaic sequences within the texts allowing the researcher to know how to sequence the presentation of the target formulaic sequences in lesson plans and prepare for explicit instruction. The goal was to teach the sequences from the texts in the order in which they appear naturally within the actual text; consequently, the information that was obtained through Microsoft Windows was of great value to the research.

The searches that employed Microsoft Windows 2007 were accompanied by frequent breaks and repeated three times over the period of 1 week to assure the reliability of the results. Through employment of Text Lex Compare (Cobb, 2010) and Microsoft Windows Version 2007, two lists of the AFL target formulaic sequences were compiled: a subset of 81 of the Core AFL (see Appendix A) and b) a subset of 46 of the Written AFL (see Appendix B). These formulaic sequences were used in instrument design procedures and also explicitly taught to the participants in the experimental group over a period of 8 weeks.

In addition to compiling the lists of AFL formulaic sequences identified in the readings for the writing course, a list of formulaic sequences used in topic discussion in the course readings was also created (see Appendix C). These topic-induced formulaic sequences can be located by using two corpus-based techniques: KeyWords extractor v.1

(2007) and N-gram Phrase Extractor v.4 (Cobb, 2010). Both programs have been adapted by Cobb for web manipulation and are available at no cost on his website (see Compleat Lexical Tutor [www.lexitutor.com](http://www.lexitutor.com)).

KeyWords Extractor v.1 (2007) is lexical computer software used to identify single words that with unusual frequency appear in a text when compared to a reference. The program calculates the word frequencies on a per million word basis and uses the Brown corpus, a corpus of one million words of American English, as a reference. The other program, the N-gram Phrase Extractor<sup>3</sup> (Cobb, 2010), generates a list of formulaic sequences occurring with the frequency of two and higher in the texts under investigation.

The process of identifying topic-induced formulaic sequences began with the identification of single key words performed by the KeyWords Extractor v.1 (2007) program. The readings that had been previously scanned and saved as text files were copied from the files, pasted in the window provided, and submitted to the program for analysis. The program generated a list of words judged by the program as keywords in the texts under investigation.

Then, the same readings were also submitted to the N-gram Phrase Extractor (Cobb, 2010) by copying the text file in the appropriate window, selecting the number of words in a string for program to identify, and selecting the submit option. The program generated a list of formulaic sequences and offered them in a split screen (see Figure 1).

---

<sup>3</sup> The N-gram Phrase Extractor (Cobb, 2010) program has been selected as a corpus-based technique for the study based on the evaluation by Ari (2006). The investigator compared three software programs KfNgram 1.2.03, Wordsmith Tools, and N-gram Phrase Extractor in terms of user-friendliness and efficacy for searching of the lexical bundles. Among the three programs, N-gram phrase extractor was rated as the most user-friendly and able to provide useful information on raw frequencies of lexical bundles.

PHRASE EXTRACTOR (WINDOW)

Home > [Extraktor input](#) (Use 'back' to preserve text/settings)>  
 Repeated string output 5 down to 2 word strings with freq>1 for **Gun control (3408 wds)**

**5-wd strings: 2422**  
 Repeated: 3 (0.12 %)

[DEATHS IN THE UNITED STATES](#) 2  
[FORTY-NINTH WORLD HEALTH ASSEMBLY](#) 2  
[RECENTLY](#) 2  
[THE FORTY-NINTH WORLD HEALTH ASSEMBLY](#) 2

**4-wd strings: 2625**  
 Repeated: 11 (0.42 %)

[AS SOON AS I](#) 2  
[DEATHS IN THE UNITED STATES](#) 2  
[FIREARM DEATH RATES VARY](#) 2  
[FIREARM DEATHS IN THE UNITED STATES](#) 2  
[FORTY-NINTH WORLD HEALTH ASSEMBLY](#) 2  
[IN THE UNITED STATES](#) 3  
[MORE GUNS = LESS CRIME](#) 2  
[SECOND AMENDMENT TO THE CONSTITUTION](#) 2  
[THE FORTY-NINTH WORLD HEALTH ASSEMBLY](#) 2  
[WORLD HEALTH ASSEMBLY](#) 2  
[WORLDWIDE PUBLIC HEALTH PROBLEM](#) 2

**2422 total 5-wd strings:**

001 1 Laws, has analyzed crime rates in the 10 STATES THAT PASSED RIGHT-TO-CARRY laws from 1977 to  
 002 ff our streets and in jail. • Require a 10-YEAR SENTENCE FOR USING KIDS in a trafficking crime.  
 003 aths of 39, 595 people in U.S. 15.6 per 100,000 , MAKING FIREARMS INJURIES the seventh leading  
 004 m deaths in the United States 14.24 per 100,000 EXCEEDS THAT OF ITS economic counterparts 1.76  
 005 ata. We compared age-adjusted rates per 100,000 FOR EACH COUNTRY AND pooled rates by income gro  
 006 areas. With each felon committing over 187 FELONIES PER YEAR, THE \$430,000 it costs to let one  
 007 he millions who have succumbed: between 1983 AND 1986, THERE WAS A 53 percent increase in femal  
 008 ns who have succumbed: between 1983 and 1986, THERE WAS A 53 percent increase in female gun-own  
 009 loped countries, epidemiology. . . . In 1990, SELF-DIRECTED AND INTERPERSONAL VIOLENCE caused 2  
 010 61% of suicides and firearm-related. In 1993, A FIREARM WAS INVOLVED in the deaths of 39, 595 p  
 011 ple and cut crime? A study published in 1995 SHOWED THAT GUNS WERE used defensively about 2.5 m  
 012 delphia passed a concealed-carry law in 1995, THE NUMBER OF PEOPLE with such permits rose by 10  
 013 at in concealed-carry states only about 2% OF PEOPLE HAVE EVEN bothered to get a permit, and th  
 014 dle. I grabbed the collars of Jack, our 200-POUND IRISH WOLFHOUD, AND HIS 140-pound malamute b  
 015 one divorce later, I headed for with my 3-YEAR-OLD SON, JORDAN THE DOGS since departed . When I  
 016 related deaths in the United States and 35 OTHER HIGH- AND UPPER-MIDDLE-INCOME countries Divisi  
 017 , we describe firearm-related deaths in 36 COUNTRIES AND ADDRESS SEVERAL questions: how do fire  
 018 methods combined. Thirty-six 78% of the 46 COUNTRIES PROVIDED COMPLETE DATA. We compared age-ad  
 019 5 million times a year and that in only 5 PERCENT OF CASES WERE defenders harmed after they bra  
 020 bed: between 1983 and 1986, there was a 53 PERCENT INCREASE IN FEMALE gun-owners n the U.S. - f  
 021 ited States, where 71% of homicides and 61% OF SUICIDES AND FIREARM-RELATED. In 1993, a firearm  
 022

http://conic.textutor.ca/hupiles/eng/users/Gun%20control%20059\_02.html#FORTY-NINTH WORLD HEALTH ASSEMBLY RECENTLY

Internet | Protected Mode On

160%

Figure 1. N-gram Phrase Extractor output for the selected readings on gun control.

The left narrow window of the split screen, shows the strings of words with frequencies of two and higher as they appear in the text. The large window to the right, shows the sequences of frequency of one and higher and the contexts in which they appear.

The words identified through KeyWords Extractor v. 1 (2007) were manually compared to the formulaic sequences produced by N-gram Phrase Extractor (Cobb, 2010) analysis because there were several instances in which the words of high range were not found in the list generated by the N-gram Phrase Extractor program. Also, it was important to select the formulaic sequences that could be used to support a position, as well as those that could be used to refute it. The subsequent manual investigation yielded additional formulaic sequences that were included in the list of formulaic sequences on topic discussion and added to the materials for explicit instruction. There were 30 topic-induced formulaic sequences equally distributed between the three themes examined in the writing course.

## **Participants**

Participants were students from five high-intermediate writing classes in an Intensive English Program (IEP) at an R1 university<sup>4</sup> in the United States. The students came from various language backgrounds (i.e., Arabic, Bambara, French, Japanese, Korean, Mandarin, Portuguese, Russian Spanish, Thai, and Turkish) all have taken a standardized English proficiency placement exam for the IEP. Some directly placed in the high-intermediate level class and some moved from the intermediate level to the high-

---

<sup>4</sup> An R1 designator for U.S. universities is used by the Carnegie Classification of Institutions of Higher Education to indicate those universities that engage in extensive research activity. The classification system includes the following: 1) Offer a full range of baccalaureate programs; 2) Are committed to graduate education through the doctorate; 3) Give high priority to research; 4) Award 50 or more doctoral degrees each year; 4) Receive annually \$40 million or more in federal support.

intermediate level after successfully passing the final exams in the previous level. The learners in both groups completed a background survey (see Appendix D) and gave permission to the researcher to allow her access to their written assignments (i.e., C-tests and essays) they completed as a regular course component. There were 64 learners who signed the consent form and gave permission to have their work used as data in the present study. Participants with the writing class attendance lower than 75% were excluded from the study because their absence from class could have had a negative effect on their performance on the tasks used as elicitation methods for the present research. As a result, there were 63 learners who participated in the study. Out of 63 participants only 44, that is, 17 from the control and 27 from the experimental group, completed all of the study tasks and were used in the data analysis.

### **Teacher**

The course instructor for the control and experimental groups was different from the researcher. At the time the study was conducted, she was 32 years old, was completing her doctorate in Applied Linguistics at the University of Utah, and teaching classes at three different schools. She obtained her Master of Arts in English/Linguistics at the University of Warsaw, Poland and had been teaching English as a second language in the United States for 9 years. She had experience teaching courses in all four skill areas (i. e., grammar, listening/speaking; reading, and writing) at various levels of language proficiency. Prior to her participation in the present study, the teacher had taught the intermediate writing class at the IEP at the University of Utah for three terms and had her syllabus laid out. The approach she had to teaching vocabulary in a writing

class was probably similar to the strategies most frequently uses in vocabulary instruction in a writing class. It consisted of having students complete the vocabulary building activities in the writing textbooks adopted for a given writing course, addressing students' questions about the vocabulary as they arose in class discussions, and advising students to look up the new words in a dictionary.

The teacher had no previous training on the teaching of formulaic language. The texts with the formulaic sequences that were typologically marked were given to the teacher to distribute to the students, and her involvement consisted of conducting the activities in a sequence provided by the researcher, but she selected when the activities would fit best in her writing syllabus. The activities with directions and answer keys were all given to the teacher to follow. The teacher was asked to read the directions carefully so that they could be communicated to the students but she was never instructed to complete the activities herself; though, it is likely that she superficially reviewed the questions to get an idea of what they are about. She was instructed to address students' questions as they arose.

The teacher provided feedback on the students' essays in both control and experimental groups. To control for the possible effects of the written feedback on students' vocabulary use in essays, the teacher was instructed to provide feedback as if the students were receiving no instruction of formulaic sequences. Given how busy the teacher, a graduate student teaching multiple classes and studying for her comprehensive exams, it is highly unlikely that the teacher had time to invest time in providing feedback to the experimental group in a way dissimilar from the feedback she gave to the control group.

### **Instruments for Quantitative Data Elicitation and Evaluation**

Supplied with the information regarding which Core AFL, Written AFL, and topic- induced formulaic sequences occurred in the readings for the writing course, the researcher developed the instruments for quantitative data elicitation. Following this, the scoring guides used in the present study were developed.

In order to elicit productive knowledge of the target formulaic sequences in a controlled situation and answer Research Question 1 (i.e., Is there a significant difference in students' abilities to *produce*, in a controlled situation, formulaic sequences frequent in academic prose (i.e., selected AFL [Simpson-Vlach & Ellis, 2010] between the students who receive explicit instruction and those who do not?) and Research Question 2 (Is there a significant difference in students' abilities to *produce*, in a controlled situation, formulaic sequences in topic discussion between the students who receive explicit instruction and those who do not?) tasks in a form of a C-test were developed. The C-test on the target AFL sequences is offered in Appendix E, and the C-test on the target topic-induced sequences is offered in Appendix F. Both C-tests were presented in the format used in Jones and Haywood (2004) such as "*Beer suspected that too much of th\_\_\_ ki\_\_\_ o\_\_\_ chemical might encourage . . .*" (p.279).

The C-test for assessing the students' ability to produce the target academic formulaic sequences was developed by using the Corpus of Contemporary American English (COCA, Davies, 2008), the largest corpus of American English currently available. COCA was selected as a reference corpus primarily because it provided the researcher with a plethora of examples of contexts in which AFL formulaic sequences



occur and could be easily extracted and adapted for the use as items on a C-test.

COCA is divided into five sections by genre: spoken, fiction, magazine, newspaper, and academic. It is a dynamic corpus, updated annually and/or biannually by adding about 25 million words. COCA currently contains 425 million words equally distributed between the genres, but at the time the C-test for the present study was developed, the corpus contained 400 million words, that is, 80 million words per each of the five corpus sections. One of the distinguishing features of COCA is that it offers an on-line and free-of-charge access to the largest collection of written texts from various academic disciplines currently available; therefore, abounding with examples of contexts in which the target formulaic sequences were used.

The COCA interface allows for various searches (e.g., by word, phrase, lemma [i.e., a citation form of a word that includes all inflections of the word], . . . etc.) of the corpus in a myriad of ways (e.g., by genre, the time-period, word/word-string frequency, etc.) that follow the same path: the selection of a preferred data display (i.e., list, chart, KWIC[Key Word in Context, a method for indexing text] and compare), submission of the search string; selection of the corpus sections to be searched, and setting up the limits and data sorting preferences. (For an in-depth description of the corpus, see Davies (2011) and for a tutorial on how to conduct searches in COCA see <http://corpus.byu.edu/coca/>).

The final version of the C-test consisted of 51 passages and is offered in Appendix E. To develop the C-test, 81 formulaic sequences were first randomly selected from the list of the target AFL (Simpson-Vlach & Ellis, 2010) formulaic sequences with 27 formulaic sequences chosen from the list of the target Core and 54 from the list of the

target Written AFL formulaic sequences. Using the COCA web interface, the target formulaic sequences were first entered one at a time as search criteria and, second, the searches were restricted to the academic genre. The search results were reported in a list of sentences from various academic journals. The sentences were examined and selected for further consideration based on the perceived difficulty of the topics they discussed. Only the sentences that dealt with the topic potentially familiar to a wide student audience were examined further. By clicking on the title of the journal article, a larger window opened up to display a passage-length context. The passages were selected for the C-test according to three criteria: a) the sentence/passage had to provide enough context for the use of the target formulaic sequences and b) the level of vocabulary in the sentence/passage had to approximate that of the course readings.

The C-test was piloted twice. It was first piloted with a native speaker and a non-native speaker of high-proficiency to provide initial feedback. Based on their input, revisions were made to three passages of the C-test. The second draft of the C-test was piloted with a group of 16 ESL students enrolled in a writing course, which was later used as a research setting during the second term of the semester. The C-test was reviewed and additional changes were made to the test. The changes include the reduction of the C-test length by 30 passages, simplification of the language of the passages, and inclusion of additional content to improve the context in which target formulaic sequences appear.

The final version of the C-test used as a measure of learners' gains in production of topic-induced formulaic sequences consists of 10 passages and is presented in Appendix F. The C-test was developed by referring to COCA (Davies, 2008-) and also

Google, an online search engine. The steps in the process of creation of the C-test items on the task resembled the procedure for writing the C-test for the assessment of target AFL academic sequences. The C-test was piloted once with a group of 18 non-native English speakers of high-intermediate language proficiency. Minor changes to the contexts of two items were made based on the information collected through the pilot test.

In addition to answering the questions on students' abilities to produce the formulaic sequences in controlled situations, the present study attempted to answer the questions regarding the students' abilities to produce formulaic sequences in uncontrolled situations, namely, their own writing. To answer Research Question 3 (i.e., Is there a significant difference in students' abilities to *produce*, in an uncontrolled situation, formulaic sequences frequent in academic prose (i.e., selected AFL [Simpson-Vlach & Ellis, 2010]) between the students who receive explicit instruction and those who do not?) two multidraft argumentative essays students wrote as a part of class assignments were used: Essay #1 and Essay #3. For both compositions students wrote two drafts and revised them based on the feedback from the teacher and/or their peers. Each essay was completed over a period of up to 2 weeks.

The tasks involved in writing the two essays were similar. The students employed the same essay rhetorical structure (i.e., present two sides to an issue and defend your position); however, in each essay, they discussed a different topic. In the first essay, participants focused on considering the issues related to genetic engineering and, in the second, those involved in international adoptions. The examination of the use of formulaic sequences in two papers on different topics seems to be a practice in research of L2 learners use and acquisition of formulaic sequences (see, for example, Hyland, 2008; Granger,

1998; Jones and Haywood, 2004). To the researcher's knowledge, the effects of a topic have not been documented and the students were asked to complete the same type of assignment, that is, to argue for a position. Having said this, the researcher acknowledged that there was a possibility of task effect as the students may find it easier to discuss one topic more than the other and, thus, are able to allocate their attention in the perceptually easier topic to making their written prose more academic by employing formulaic sequences. The researcher attempted to lower the effect of attention allocation by making sure that the students had about 2 weeks to develop their ideas and work on their papers.

To examine the possible gains in learners' abilities to produce topic-induced formulaic sequences in their writing and, therefore, answer Research Question 4 (i.e., Is there a significant difference in students' abilities to *produce*, in an uncontrolled situation, formulaic sequences in topic discussion between the students who receive explicit instruction and those who do not?), students were given an in-class 40-minute argumentative essay to write (see Appendix G for directions) based on the readings they had completed in preparation for the class. The students wrote on the following topic:

*“Some people agree with Thomas Atwood, the President of America’s National Council for Adoption, who states: “National boundaries should not prevent abandoned children from having families.” Others take the position that orphaned children should remain in their home countries.*

*What is your stand on the issue of international adoption? Should a country allow international adoptions or limit adoptions to domestic adoptions only?”*

In order to answer Research Question 5 (i.e., Is there a significant difference in the overall quality of students' essays as measured by the trained essay raters' judgments

between the students who receive explicit instruction and those who do not?), the study examines the effects of explicit instruction of the formulaic sequences, both academic and topic induced, on the students' writing products and looks at the students' overall ability to write. Essay #3, the final multidraft argumentative essay students wrote for the class, was used as the product to answer this question. The in-class 40-minute essay on the same topic was used as a pretest.

In addition to designing the instruments for elicitation of quantitative data, the researcher created the rubrics for the assessment of the accuracy of the production of formulaic sequences in controlled and uncontrolled situations. The rubrics were based on the scale developed by Jones and Haywood (2004) and adapted to better fit the purposes of the present study. The rubric for measuring the production of the formulaic sequences in a controlled situation appears in Figure 2.

- 3 - correct phrase; spelling issues possible but do not overlap with the issues with inflectional and/or derivational affixation;
- 2 - correct phrase; problems with inflectional morphology (e.g., *in term of* instead of *in terms of*) and issues with demonstrative pronouns (*in that case* instead of *in this case*);
- 1 - incorrect phrase but an attempt at production of correct phrase evident, which can be described as one of the following:
  - a) Problems with derivational morphology (e.g., *slaughtery* instead of *slaughter*)
  - b) Substitution of no more than ONE word within a phrase with another word of the same word category that is very similar in spelling, pronunciation, and/or meaning (e.g., *the effects of* instead *the efforts of*; *in the case* instead of *in this case*)
- 0 - little to no attempt to complete the phrase OR the use of any combination of the issues described under the rating of 1.

**Figure 2.** Description of the scale for measuring production of formulaic sequences on a C-test.

For the assessment of the production of formulaic sequence in uncontrolled situations (i.e., the essays), a more conservative rubric than the one employed for the scoring of C-tests was developed in Figure 3 primarily to assure the reliability of the results. To anticipate one-word substitutions within a phrase with another word of the same word category, that is similar in spelling, pronunciation, and/or meaning in the phrases students produced in their essays, would have been very difficult to do, and more importantly, the process of evaluations would have been very difficult to carry out as it would create an extremely large pool of searchable phrases in the essays. As can be seen from the rubric below, only the phrases with substitutions in preposition and pronoun, provided the replacements are the same part of speech, were considered as attempts at production of the target phrase in writing. In addition, the formulaic sequences that had any modifications other than a pronoun or preposition substitution were not considered as attempts at production of the target formulaic sequences. The rubric follows:

- 3 - correct phrase; spelling issues possible but cannot be mistaken for the issues with inflectional and/or derivational affixation;
- 2 - correct phrase; problems with inflectional morphology (e.g., *in term of* instead of *in terms of*)
- 1 - incorrect phrase but an attempt at production of correct phrase evident which can be described as one of the following:
  - a) Substitution of a preposition (e.g., *in the other hand* instead of *on the other hand*)
  - b) Pronoun confusion (e.g., *his or her* instead of *him or her*) or reversed order in compound constructions (*her or his* instead of *his or her*)
  - c) Omission of a function word inside the phrase (e.g., *as result* instead of *as a result*)
- 0 - no attempt to produce a target phrase OR any combination of the issues described under the rating of 1.

**Figure 3.** Description of the scale for measuring the production of formulaic sequences in writing.

To assess the writing proficiency of the students at the end of the instruction, a writing rubric was used. The rubric is offered in Appendix H and represents a modified version of the rubric developed by Jacobs, Harfield, Hughey, and Wormeth (1981). The scoring guide was used to assess the 40-minute diagnostic essays and the multidraft essays on the same topic. The scoring guide is based on a four-point scale, and student essays were assigned a score on the scale from 1 to 4 for each of the following five aspects of writing: content, organization, vocabulary, language use, and mechanics. In the scale, content refers to the development of the thesis and amount of support provided; organizations refers to the unity of ideas; vocabulary refers to clarity of expression and range of vocabulary used; language use refers to syntax complexity and grammar accuracy; and mechanics to spelling, punctuation, capitalization, and general formatting. The scores were weighted by multiplying the score in the following manner: content by 7.25; organization by 5; vocabulary by 5; language use by 6.25; and mechanics by 1.25. After the ratings were translated into scores, they were combined to create an overall score for each essay. After the target formulaic sequences were identified in the readings, the instruments for quantitative data elicitation created, and assessment tools drafted, the activities used for explicit instruction of the formulaic sequences were created.

### **Description of the Treatment**

The treatment was provided over a period of eight weeks or one academic term in the IEP in a high-intermediate writing class designed to prepare students for the study at a university. The treatment consisted of teaching the formulaic sequences from the AFL list (Simpson-Vlach & Ellis, 2010) and those used in the discussion of the topics located

in the reading materials for the writing course. The students were told that the class would focus on teaching academic vocabulary and that learning the phrases could help them improve their writing of academic argumentative essays on selected topics. Explicit instruction of the target formulaic sequences was integrated into the lesson plans for the writing class and provided in increments of 5-20 minutes per class in at least two out of the 4 days of instruction per week during an 8-week term. Explicit instruction of the target formulaic sequences was carried out alongside activities for implicit vocabulary learning (e.g., Hujistin, 2001; Schmitt, 2000), provided in meaningful contexts, and aligned with the three psychological processes necessary for successful vocabulary learning--noticing, producing, and generating (Nation, 2001). Students were engaged in activities that focused on the development of receptive knowledge and then on the productive knowledge of the target formulaic sequences.

The instruction of the target formulaic sequences was distributed over a period of 8 weeks and followed the order in which the readings on a given topic were discussed. To illustrate, the academic formulaic sequences “*whether or not*” and “*are likely to*” were used in the reading on genetic engineering that was taught first; whereas, “*his or her*” and “*the value of the*” were used in the reading on international adoptions; therefore, they were taught in the final weeks of instruction. Likewise, the formulaic sequences “*threatened species*” and “*wolf reintroduction programs*” were taught at the time the topic of grey wolf reintroduction to the wild was discussed; whereas, the formulaic sequences “*inter-country adoption*” and “*adoption agency*” were taught at the time the topic of international adoptions was examined.

The treatment proceeded in the following manner: First, the learners received the



course materials in which the target AFL formulaic sequences were framed and in which the topic-induced formulaic sequences were bolded to increase their salience to facilitate noticing. The readings were usually completed outside of class and then used in schema-building activities to help prepare students to write on the given topics. After the students became familiar with the content of the readings, they were explicitly told that the marked formulaic sequences were phrases frequently used in academic writing and that the bolded phrases were important in effective discussions of the topics examined in the writing class.

The instruction of the academic formulaic sequences proceeded in the following manner: The students were told that the formulaic sequences used in academic writing had a specific function in the text. Following this, the students were engaged in activities that would help them know more about the target formulaic sequences to encourage deep level processing to enhance the acquisition. Specifically, they were asked to examine the functions of the formulaic sequences in the context to classify the formulaic sequences based on their pragmatic function in the text or into meaning-based groups. In this activity, the students worked in groups. They were asked to read the context for the formulaic sequences, examine the pragmatic functions of the target formulaic sequences in the reading, and categorize the formulaic sequences by their pragmatic function.

Subsequently, the students were engaged in a variety of activities: gap filling exercises where they needed to complete sentences with the formulaic sequences that were provided to them; C-test type activities in which they needed to fill in the missing parts of the target formulaic sequences; dictogloss<sup>5</sup> (Wajnryb, 1988); identification of the

---

<sup>5</sup> Dictogloss (Wajnryb, 1988) is an activity which involves students in a series of activities focusing on language development. First, the students are exposed to repeated listening and/ or reading of a text. Next, they are asked to take notes while listening/reading. Finally, the students get together in pairs and groups to compare notes and to recreate the text they were just exposed to.

target academic sequences in passages; identification of the academic phrases in a passage and asking students to produce them in an oral presentation; and examining their own essays to evaluate the use of formulaic sequences. The students were explicitly told to keep the handouts with the lists of the target academic formulaic sequences and refer to them in the process of writing. They were also reminded that they were expected to use the target formulaic sequences in their own writing.

The instruction of the topic-induced formulaic sequences continued over the period of three to four classes per topic. The students were first engaged in completing activities aimed at their ability to produce the topic-induced formulaic sequences in controlled situations and then in the activities that allowed students to produce them in their speech and writing. Over the course of the term, the students completed a variety of activities aimed at teaching the students for the purposes of writing: cloze-type activities such as matching the formulaic sequence with the context and filling in the missing parts of the formulaic sequences; a 2/1/30 activity<sup>6</sup>; a version of Jeopardy with formulaic sequences; building an argument, and asking students to examine their own writing for the use of the topic-induced formulaic sequences.

The treatment prior to the collection of data on the production of the topic-induced formulaic sequences in controlled (i.e., C-test) and in uncontrolled (i.e., an essay) situations followed the same sequences; namely, it focused first on the development of receptive knowledge and then attempted to bring the receptive knowledge to productive

---

<sup>6</sup> The 2/1/30 activity represents a modified 4/3/2 activity described by Nation and Gu (2007). The activity was carried out in the same manner as described by the authors except that the timing was adjusted from 4/3/2 minutes to 2/1/and ½ minutes respectively. The activity aims at developing learners' productive skill of speaking. It was used as precursor to the writing activity. It is designed as a pair activity. One learner is the speaker and the other is the listener. In this study both learners were given a text to read. After they have read the passage on a familiar topic, the speaker would talk to one partner in 2 minutes. Then s/he would move to the next and speak in 1 minute and finally to the third partner in 30 seconds. During this time, the listeners did not interrupt, but only listened. After that, the listeners and speakers changed roles.

use. However, the treatment incorporated a slightly different set of activities and was conducted at a different time in the semester. The treatment on topic-induced formulaic sequences was done prior to data collection on the C-test and was conducted during mid-term prior to students drafting the essay on a grey wolf as an endangered species. The treatment consisted of the following steps. The students received the readings with the target formulaic sequences bolded. After they became familiar with the content of the readings, they focused on five vocabulary exercises as devised by the teacher over the period of four classes. The first two activities were gap filling activities. The first gap filling activity was a matching activity that the students did in pairs. One student received the phrases and the other sentences. In pairs they were asked to match the phrases with the sentences. After the activity was completed, the students shared their answers with the class. For homework, the students were given the matching activity on a handout to complete individually. The second activity, a "cloze-type" activity in which the students completed the missing parts of words, was assigned for homework. The third activity was an adaptation of an American TV quiz show "Jeopardy" and completed in class. The students were separated in groups and competed against one another in answering the questions regarding the target vocabulary correctly. For each correct answer the group received a certain number of points that was determined by the difficulty of the question. The more difficult questions were worth more points, the less difficult questions were worth fewer points. The students were asked to complete a paper version of the game outside of class. The fourth activity was "Build an Argument," a writing activity requiring students to utilize a set of target formulaic sequences on the topic of discussion in an argument in favor of or against a controversial issue. The final activity required of

students to examine the first draft of their essay to find out whether or not they had utilized the topic-induced formulaic sequences in their prose.

The treatment prior to the elicitation of data on the production of the topic-induced formulaic sequences in an uncontrolled situation was conducted during the time students explored the topic of international adoptions and in the final weeks of the semester. The treatment consisted of the following: The students received the readings with topic induced formulaic sequences on the topic of international adoptions bolded. Over the period of four classes, the students completed five vocabulary activities. The first two activities were matching and "cloze-type" activities that were both completed in class and in pairs. The same activities were assigned for homework that was intended to be completed individually. The next activity was the 2/1/30 activity followed by a type of a cloze test where students had to fill in the blanks. For this gap filling activity students were not provided with the target forms to match nor were they offered the beginning segments of the words. The final activity was a writing activity requiring students to utilize a selected group of formulaic sequences on the topic of international adoptions in building arguments in favor of or against the controversial issue.

All of the activities as a type (e.g., matching, fill in the blanks, build an argument, etc.) used in explicit teaching of formulaic sequences were piloted (except for three—an examination of an essay draft to identify and encourage the use of formulaic sequences; 2/1/30 activity; and a sorting and categorizing activity) with a group of high-intermediate students and on the topic of the wolf as an endangered species. Based on the input received from the teacher, several of the activities were slightly modified. Specifically, the matching activity was modified from a group to a pair activity; a PowerPoint

Presentation for the Jeopardy game was prepared; and minor issues in the "cloze-type" activity were corrected.

### **Study Procedures for Quantitative Data Collection**

During the first week of classes, prior to the explicit instruction on formulaic sequences, the pretest on the production of the academic formulaic sequences in a controlled situation (C-test), was given. The pretest on the production of the topic-induced formulaic sequences in an uncontrolled situation (i.e., in-class 40 minute argumentative essay) was conducted on the first day of the second week of instruction as the course instructor devised. It was the final segment of a procedure that began by giving students four passages on the topic to read prior to class accompanied by a set of comprehension questions that were mainly included to improve the likelihood of students' reading the assigned texts prior to class. The students were further encouraged to complete the readings by being informed in the directions for homework that their preparation for the class was expected and that their success in class was dependent on their completion of homework assignments. On the day of the pretest, the learners were asked to write an impromptu in-class argumentative essay (for directions see Appendix I).

The same tests were given as posttests during the final week of the course. The post C-test was completed on the last day of classes. As for the in-class essay, both groups of students were given the readings with the comprehension questions to complete and the same instructions as those in the pre-test. Over the period of three classes, the control group engaged in writing-related activities inside and outside of class while the experimental group received explicit instruction on formulaic sequences consisting of

activities that were designed to raise learners' awareness of formulaic sequences, provide opportunities for students to manipulate them in controlled situations, and finally use them productively in speech and writing. On the fourth day, both groups were asked to write an impromptu in-class essay.

The posttest on the production of topic-induced formulaic sequences in a controlled situation (i.e., C-test) was administered at the beginning of Week 6 of the term, after the students in the experimental group had become familiar with the activities used in the present study for explicit instruction of formulaic sequences. By so doing, the researcher attempted to allow learners in the group to allocate their attention to completing the activity by focusing on the language, rather than, possibly, focusing on the activity procedures, such as on figuring out what was involved in the next step of the activity. The final score given to the accuracy of use of formulaic sequences in a C-test is a sum of the scores given to each formulaic sequence in the tests.

The second set of data was collected through the tasks designed to measure the production of academic formulaic sequences in an uncontrolled situation. To assess whether students' production of the selected AFL (Simpson-Vlach & Ellis, 2010) formulaic sequences had developed over the period of the study, two essays that participants wrote were examined: multidraft Essay 1 and multidraft Essay 3. The students started working on the first composition during the second week of classes, and on the third composition during the seventh week of classes. Including the days students spent discussing the topics of their essays in class, the students had a little over two weeks to complete each of their essays.

### Study Procedures for Quantitative Data Analysis

The data for the study included scores students received on the following:

- pre- and posttests on the production of academic formulaic sequences in a controlled task (C-test on the target AFL[Simpson-Vlach & Ellis, 2010] formulaic sequences);
- test on the production of topic-induced formulaic sequences in a controlled task (i.e., C-test on the target topic-induced formulaic sequences)
- pre- and posttests on the production of topic-induced formulaic sequences in an uncontrolled task (i.e., in-class 40-minute argumentative essays)
- Essay #1 and Essay # 3 as measures of the production of academic formulaic sequences in an uncontrolled task;
- Pretest (i.e., in-class 40-minute argumentative essay) and posttest (i.e., Essay #3) of the learners' writing ability.

All of the C-tests were collected on the days they were administered. They were evaluated by the researcher and a trained rater who both used the *Scoring Rubric for Production of Formulaic Sequences on a C-test*. In case of discrepancies between the scores, which were few, the raters re-rated the tests. It was found that the score discrepancies were mainly due to technical errors in recording the scores. The raters reached the interrater reliability of 1.00. The final score given to the accuracy of use of formulaic sequence on the C-test is the sum of scores given to each formulaic sequence.

Before the essays used as measures of students' abilities to produce academic formulaic sequences in their writing could be assigned a score, lexical analyses had to be

conducted to extract the target AFL formulaic sequences. The students submitted their multidraft essays (i.e., Essay #1 and Essay #3) for evaluation in an electronic format via email to their instructor who forwarded the essays to the researcher. In preparation for examination of the students' essay, the researcher coded, mixed, and imported the essays written by experimental and control groups in one large word document with the codes written on top of each essay.

The procedure for identification of the target academic formulaic sequences in the participants' essays greatly resembled the procedure for the detection of the formulaic sequences of the AFL (Simpson-Vlach & Ellis, 2010) in the course readings in terms of the steps involved and the two computer software used: a) Text-Lex Compare v.2. 2 (Cobb, 2010) and b) Microsoft Windows version 2007.

The target academic formulaic sequences and participants' compositions were submitted, one at the time, to the Text-Lex Compare v.2. 2 (Cobb, 2010) via Method 1 as follows: the document with the participants' compositions were entered in the window entitled "Old Text," and the target Written and Core formulaic sequences were submitted separately as data in the window entitled "New Text". The Text-Lex Compare program generated four lists of phrases: "Unique to old," a list of phrases that appear in the "Old" text or the students' essays; "Unique to new", a list of phrases exclusive to the "new", that is, the target Core AFL or the target Written AFL arranged by frequency and alpha listing; and "Shared," a list of phrases that appear in both the readings and the examined target subsets of the AFL formulaic sequences.

The information generated by the program in the list entitled "Shared" was copied in a Word document and saved for the use in further analysis. The Word document



contained a table that the researcher created to sort the results of the Text-Lex Compare program and insert notes. The table was organized with rows that represented each participant separately and columns that represented the Text-Lex Compare program analysis with results arranged by participants.

The Microsoft Word version 2007 program with its search feature “Find” was used in the process of locating the formulaic sequences in the participants’ compositions. The document that contained all of the participants essays coded was submitted to analysis. The search was conducted by activating the “Find” feature of the program and by entering, the formulaic sequences offered in the "Shared" list that had been previously generated by the Text-Lex Compare v.2. 2 (Cobb, 2010). These sequences were used as the search criteria and dealt with one at a time. The program highlighted the target formulaic sequence in the text under examination. Each time the target formulaic sequence was located, the researcher examined the form and function of the formulaic sequence to determine whether a) the forms and pragmatic functions of the formulaic sequences highlighted in the text matched the form and function of the target formulaic sequences; b) the formulaic sequences were a part of students’ prose or the quoted and/or unquoted reference materials; c) there were instances of an overlap of two or more formulaic sequences. The researcher bolded all of the formulaic sequences in the document and recorded her notes in the table along with the results of the Text-Lex Compare program.

After the formulaic sequences identified by the Text-Lex Compare program were located and marked in bold in the text, the researcher continued the examination of the compositions using the Microsoft Word program and its feature "Find" to locate possible

erroneous target formulaic sequences (e.g., issues with spelling, problems with morphology, dropped words within the formulaic sequences). The search was conducted by entering partially realized forms of the target formulaic sequences as search criteria. For example, when the essays were examined for the occurrences of "*in other words*," the following criteria were submitted in the search "*in other*" first and then "*word*." If a word or a word string that met the search criteria was found in the text, the researcher examined the context in which it appeared to determine whether or not it was element of a formulaic sequence. If the researcher determined that it was a formulaic sequence bolded it; however, if they were determined to be random strings of words, they were left unmarked. For example, the search for the element "word" resulted in identification of the following nontarget-like formulaic sequences: "*in another words*" and "*in other word*." After the examination of the context and the pragmatic function assigned to the two strings of words, the researcher determined that both represent attempts at the production of the target formulaic sequence "*in other words*" and not some random word combination. For this reason, the phrases were made salient by bolding to be easily located in the text for future examination. The process of identification of the target formulaic sequences in the students' compositions was repeated three times over a period of four days to assure the reliability of scoring of data. The researcher took 15- to 30-minute breaks between searches after every 30 formulaic sequences.

After the target formulaic sequences were located and made salient in the students' compositions, the researcher reviewed the compositions with bolded text. Before the formulaic sequences could be evaluated using the scoring guides, the researcher needed to: a) eliminate from the analysis the formulaic sequences that were

not a part of students' prose but were from the material borrowed from sources to support the student-generated text; b) identify multiple occurrences of the same formulaic sequences to assign one score to the formulaic sequence if they were consistently produced as target-like forms or calculate an average of scores assigned to each occurrence using the scale for evaluation formulaic sequences in writing; and c) consider the instances of an overlap of two or more formulaic sequences such as "*the same time*" and "*at the same time*" to evaluate them as one and not two formulaic sequences so not to inflate students' scores. After the researcher identified the formulaic sequences that were a part of students' prose, single occurrences of the target forms, and forms that overlap, the researcher evaluated the formulaic sequences using *the Scoring Rubric for Production of Formulaic Sequences in Writing*. The final score given to the accuracy of the form of the formulaic sequences in the essay was a sum of the scores given to each formulaic sequence in the tests.

To examine the learners' abilities to produce the formulaic sequences on topic discussion in writing, a 40-minute in-class essay on the topic of international adoption was used as a pre- and posttest. Both the pre- and posttest essays were collected from the students, typed, and saved on a computer. An independent evaluator and the researcher compared the electronic versions of the essays to the handwritten essays to ascertain that they were entered correctly. The examination of the typed and handwritten essays revealed some minor inconsistencies. These inconsistencies were corrected so that the essays used in the subsequent analyses accurately represented the content of the handwritten in-class essays.

The researcher conducted a lexical analysis of the pre and posttest essays to

extract the target topic-induced formulaic sequences. The procedures for identification and evaluation of the topic-induced sequences in students' writing were the same as procedures for the identification and scoring of the AFL formulaic sequences (see pp. 67-72) with the exception that the targets of the search were not the AFL but the topic-induced formulaic sequences on the topic of the wolf as endangered species.

To measure the overall effects of explicit instruction on the ESL learners, the scores that represented the average of scores assigned to each essay by trained raters were used. As noted previously, the raters used a modified version of the ESL Composition Profile rubric (Jacobs et al., 1981). The raters were either current or past L2 writing teachers. Both had experience in grading student essays using an analytic scale and both participated in the process of the modification of the Jacobs et al. scale. The raters practiced using the rating rubric on seven 40-minute pretest or final essays written by the students in the study. These essays were rated by each rater and the researcher, and then the scores were compared and discussed. If there was a difference in scoring, the reasons for choosing the ratings were discussed until all the raters agreed on a given score. In the first round, the raters rated within one point of each other on all essays on a 4-point scale. They commented that they understood the five criteria and felt comfortable using the whole range of the scale.

All the essays were graded by two raters and the scores were averaged. The interrater reliability for the overall essay scores was .83. For the four aspects of writing the interrater reliability was the following: .82 for content, .64 for organization, .66 for vocabulary, .54 for language use, and .45 for mechanics. If the first two raters disagreed by more than one point in any of the seven criteria on a given essay, the researcher acted

as a third grader who also graded the portions of the essay in question before the scores for the essay were averaged. There were four essays that the researcher rated to resolve the dispute between the scores assigned by the two raters. The raters were asked to assign individual scores to each student on each of the criteria on the rubric. The researcher translated the rating into weighted scores and entered the scores into an Excel spreadsheet on the computer. The scores on each criterion were then added obtain an overall score for the essay. The raw data score was used as the score for overall essay quality.

In addition to answering Research Questions 1-5 that focused on quantitative data, the study attempted to glean insight into the approaches ESL learners take in learning to identify and produce formulaic sequences in their writing. The approach taken to collect and analyze qualitative data is offered in the next section.

### **Study Procedures for Qualitative Data Elicitation**

Although the primary aim of the present study was to investigate the effects of explicit instruction of formulaic sequences on L2 writers, the researcher realizes that not all research questions can be answered using quantitative data analyses. While the effect of treatment on students' abilities to produce the target formulaic sequences on a test and in writing was well as on students' writing abilities can be examined by quantitative analyses, other aspects of the topic under investigation, such as the use of learner strategies, seem to be better examined using qualitative research methods (Davis, 1995). In order to better understand the effects of explicit instruction of formulaic sequences on L2 writers, it is necessary to bring in the learners' perspectives on the outcomes of this

learning through a qualitative data elicitation technique, one of which is an interview. Interviews represent one of the most frequent research techniques in qualitative investigations (Davis, 1995; Nunan, 1992). They are broadly defined as structured exchanges of information with an informant and have gathering information as a primary goal. As a research tool, interviews have been viewed on a continuum with varying degrees of structure (Nunan, 1992) where the structured and non-structured interviews lie on the extremes of this continuum.

Jones and Haywood (2004) had intended to conduct individual interviews with the three participants in the study, two judged stronger and one weaker in terms of vocabulary knowledge that based on their performance on the pre tests; however, due to the time constraints, they had to resort to using surveys. The surveys questions asked for information on student background in terms of their educational and formal experiences with English and also on vocabulary learning strategies. The three students reported on using strategies in learning the formulaic sequences, such as repeating them, noting them in a vocabulary workbook, or creating sentences; however, in an examination of the reports provided by the researchers, it is difficult to pinpoint the specific characteristics associated with individual strategies that L2 writers employed that distinguish the more successful learners of the formulaic sequences from the less successful learners. Because it is important to identify which strategies assist students in becoming successful learners of formulaic sequences, the interview process that was employed in the present study attempted to glean insight into what practices set apart learners who can use the formulaic sequences and use them successfully in their writing from those who cannot.

Besides completing the tasks that allow for qualitative data analyses, a subgroup

of four participants were invited to take part in posttreatment semistructured interviews that would elicit data for qualitative analysis. The informants were chosen on the basis of their abilities to produce the formulaic sequences in controlled situations as indicated by their performance on the tasks. To pinpoint the distinguishing differences in approaches to learning of formulaic sequences for the purposes of writing, there were two learners selected from high-achieving and two from low-achieving group of learners.

The interviews in the present study were semistructured, that is, the wording and sequence of the interview questions remained the same for each informant; however, probes were used to elicit additional information as the need arose. To prepare for the interviews, the researcher crafted an interview guide to direct the interview. The interviews incorporated retrospective verbal report protocols, a research technique that requires participants to recall and report verbally strategies they use in completing a task after the event has occurred. To uncover the strategies learners applied in formulaic sequence production tasks, the researcher identified the segments in the production tasks in the pre- and posttests of each of the informants to provide a context and to help informants remember the thoughts and strategies they used while completing the task.

The interview questions consisted of experience, opinion, and knowledge questions. The interview began with the researcher introducing herself, giving the information on the time and day of the interview, and the purpose of the interview. The questions for retrospective verbal report protocols revolved around the questions listed below but varied depending on the context:

1. How familiar were you with the phrase X at the beginning of the term?
2. This is your test form the beginning of the semester. I would like you to

examine this sentence and the words that you were asked to complete it (the researcher will point to a sentence and the target partially realized formulaic sequence on a C-test). Could you please tell me what you were thinking so that I can understand how you were trying to fill in the missing parts of words (the researcher will wait for the informant to respond).

(After the reply, the researcher continues with the follow up questions) Here is the same problem on a test at the end of the semester. Could you please tell me what you were thinking so that I can understand how you were trying to fill in the missing parts of words in this test.

3. In your first draft of this essay you used the phrase taught in class. How did you go about using them this early in writing process?

4. In your first draft of this essay you did not use the phrase taught in class. In your final draft you used the phrases taught in class. How did you go about incorporating the phrases in the final draft of your essay?

After the information on the retrospective verbal report protocols was collected, the interview continued by addressing the following questions which were asked in the order listed below:

1. How do you go about learning phrases for the purposes of writing?
2. How do you go about using phrases in your writing?
3. In the writing class, your teacher used many different activities to help students learn the phrases. In your opinion, which of these activities helped you learn the phrases the best?
4. Which of the activities were not helpful to you?



5. What do you think could have helped you more to learn the phrases in this class?

The interview closed with the following question: Is there anything else you would like to share with me as it relates to your learning of the phrases in your class?

The interviews were tape recorded. The researcher listened to the information as many times as was necessary in order to represent the information accurately and take notes while listening. The notes helped the researcher to locate quickly the most relevant sections of the interview because searching the tapes for specific details may become time consuming (Patten, 1990). The researcher analyzed the data from the interview from within a postpositivist paradigm (Hatch, 2002) (i.e., ontological and epistemological viewpoint that reality can only approximated) by looking for patterns and relationships in the data with an attempt to describe the distinguishing characteristics of a selected group of informants. More precisely stated, the researcher looked for patterns in the responses of the informants in an attempt to uncover the strategies that had been used by more successful and those used by less successful learners, capture the learners' perspective on explicit instruction of formulaic sequences in an L2 writing class, and describe the process of learning to identify and produce formulaic sequences.

## CHAPTER 4

### RESULTS

In this chapter the results of the quantitative analyses will be presented for the control and experimental groups. Performance differences for each group will then be reported. These analyses will be followed by a report of the results of the qualitative analyses from the follow-up interviews.

#### Descriptive Statistics

The means and standard deviations for the scores on the tests used in the study appear in Table 1. The students in the experimental group scored higher than the students in the control group on all but one measure: the production of academic formulaic sequences in an uncontrolled situation (i.e., essay). In fact, the mean scores on the test of students' abilities to produce the academic formulaic sequences in a controlled situation and the topic-induced formulaic sequences in an uncontrolled situation for the experimental group were twice that of the control group and also more than three times higher on the test of the students' abilities to produce the topic induced formulaic sequences in a controlled situation. The mean scores on the posttest of the students' writing abilities were slightly higher for the experimental  $M=78.81$  ( $SD= 9.76$ ) than the control group  $M= 78.50$  ( $SD=9.90$ ). More importantly, the gains in writing ability for the

Table 1.

*Minimum, Maximum, Mean Scores and Standard Deviations for all Measures by Group*

Measure	<i>n</i>	Min.	Max.	<i>M</i>	<i>SD</i>	<i>N</i>	Min.	Max.	<i>M</i>	<i>SD</i>
Control						Experimental				
C-test AFS										
Pretest	24	1.00	59.00	31.00	15.04	36	4.00	92.00	31.86	18.16
Posttest	23	12.00	95.00	44.65	19.13	34	12.00	148.00	<b>81.47</b>	33.39
Essay AFS										
Essay 1	21	3.00	24.00	14.12	6.36	34	3.00	39.00	17.30	8.93
Essay 3	20	3.00	21.00	11.18	5.76	31	.00	24.00	8.77	6.66
C-test T-IFS	23	.00	24.00	7.00	6.30	35	7.00	30.00	<b>21.49</b>	7.51
Essay T-IFS										
Pretest	24	.00	8.30	3.98	2.00	39	.00	9.00	3.89	2.03
Posttest	19	3.00	10.00	4.84	2.26	35	.00	21.00	<b>8.71</b>	5.40
Writing										
Pretest	24	<b>41.25</b>	66.88	52.73	6.99	38	<b>31.00</b>	66.25	52.29	8.76
Posttest	20	<b>58.13</b>	100.00	78.50	9.91	31	<b>55.00</b>	100.00	<b>78.81</b>	9.76

Note. C-test AFS = production of academic formulaic sequences in a controlled situation (i.e., C-test); Essay AFS = production of academic formulaic sequences in an uncontrolled situation (i.e., essay); C-test T-IFS = production of topic-induced formulaic sequences in a controlled situation (i.e., C-test); Essay T-IFS = production of topic-induced formulaic sequences in an uncontrolled situation (i.e., essay); Writing = in-class essay

low performing students were larger for those in the experimental group than those in the control. The minimum score in control group on the test of the students' writing abilities was 41.25 at the beginning of the semester and 58.13 at the end of the semester with a gain of 16.88; however, the minimum score in experimental group was 31 on the pretest and 55 on the posttest with the gain of 24. The students in the control group performed higher on the measure of students' abilities to produce academic formulaic sequences in an uncontrolled situation with the mean for the control group 11.18 ( $SD=5.76$ ) and the experimental 8.77 ( $SD=6.66$ ). The mean scores on the individual tasks increased from the beginning to the end of the semester for the experimental and control groups with the exception of one: the production of the academic formulaic sequences in an uncontrolled situation, namely, an essay. The decrease in the performance on the task for the control group was of 2.94 moving from 14.12 ( $SD=6.36$ ) on Essay 1 to 11.12 ( $SD=6.36$ ) and for the experimental group of 8.52 with the mean score of 17.31 ( $SD=8.93$ ) on Essay 1 and the mean score of 8.78 ( $SD=6.66$ ) on Essay 3.

The means and standard deviations for the pre- and posttest scores on each of the five aspects of writing (i.e., content, organization, vocabulary, language use, and mechanics) are offered in Table 2. The score means on the posttest indicate that the experimental group outperformed the control on the vocabulary and language use areas of writing, but that the control group was superior to the experimental on the content, organization, and mechanics aspects of writing. In addition, despite the fact that the experimental group started lower on vocabulary aspect of writing, the group made up for the initial difference and managed to outperform the control at the end of the semester.

Table2.

*Pretest and posttest Mean Scores and Standard Deviations for the Control and Experimental Groups*

	Control		Experimental	
	Pretest	Posttest	Pretest	Posttest
	(n=24)	(n=20)	(n=38)	(n=31)
Content	11.72 (3.88)	24.94 (5.48)	10.95 (3.30)	23.83 (3.43)
Organization	10.83 (2.17)	15.63 (2.55)	11.25 (2.89)	15.08 (2.46)
Vocabulary	11.15 (1.47)	14.63 (2.01)	<b>10.99 (2.57)</b>	<b>15.96 (2.39)</b>
Language use	11.71 (2.16)	18.59 (3.12)	15.13 (2.12)	<b>19.35 (2.84)</b>
Mechanics	4.33 (.55)	4.72 (.43)	3.96 (.72)	4.57 (.49)

Note. Standard deviations given in parentheses.

### **Differences between Control and Experimental Groups**

A one-way multivariate analysis of variance (MANOVA) was conducted to investigate the differences between the control and the experimental group. MANOVA was performed on the dependent variables, which constitute students' performance in production of the academic formulaic sequences in controlled and uncontrolled situations, topic-induced formulaic sequences in controlled and uncontrolled situations, and students' writing abilities. The independent variable was group membership with 17 control group participants and 27 treatment group participants. The number of observations per student was reduced because not all participants took all the tests, but the numbers were high enough and similar enough to retain statistical power. To correct

for any possible effect of the unequal sample size, Type 1 sums of squares were calculated, which does not assume equal cell sized.

Multivariate Tests revealed that there were statistically significant differences between the treatment and control groups at posttest on the combined dependent variables,  $F(5, 38)=8.363$ ,  $p<.001$ ; Wilks' Lambda = .48; partial eta squared effect size = .524. These results indicate that there was an overall effect of group membership on academic outcome measures with students in the treatment group outperforming the students in the control group overall. The effect size of .524 is large, indicating that approximately 50% of the total variance, that is, the difference of every score from the grand mean, in posttest test scores can be accounted for by group membership.

When the univariate results for the dependent variables were considered separately, it was revealed that there were statistically significant differences between the groups on the production of academic sequences in a controlled situation  $F(1, 42)=19.80$ ,  $p<.001$ , partial eta squared = .320; of topic-induced formulaic sequences in a controlled situation,  $F(1, 42)=33.42$ ,  $p<.001$ , partial eta squared .443; and in uncontrolled situation TI,  $F(1,42)=8.661$ ,  $p<.008$ , partial eta squared .171. The groups did not differ significantly on the measures of production of academic formulaic sequences in an uncontrolled situation  $F(1, 42)=2.41$ ,  $p= .13$  and on writing ability  $F(1, 42) = .173$ ,  $p= .679$ .

Levene's test of Equality of Error Variances was significant for the measures of the students' abilities to produce the academic formulaic sequences in a controlled situation and topic-induced formulaic sequences in an uncontrolled situation, indicating that the distribution and variability of the scores for the control and experimental groups

Table 3.

*MANOVA for Control and Experimental Groups*

Measure	<i>F</i>	<i>M</i>	Partial eta square	Observed Power	Levene's Test <i>F</i>
C-test AFS Posttest	19.80**	14045.07	.320	.991	6.308**
AFS Essay 3	2.41	95.94	.054	.330	.085
C-test T-I FS Posttest	33.42**	1857.58	.443	1.00	2.444
Essay T-I FS Essay	8.66*	182.72	.171	.82	8.105*
Writing Posttest	.17	18.57	.004	.069	.061

Note: \* $p < .008$ ; \*\* $p < .001$

on these two measures were different. MANOVA is quite robust to violations to this assumption.

To ensure that the groups started out the same, the pretest data was submitted to statistical analysis. It was determined, post hoc, that submitting data to Independent T-tests on each of the pretest measures was a more appropriate measure than using a repeated measures design (e.g., MANCOVA) because homogeneity between the groups was an issue.<sup>7</sup> Because the independent T-tests on each of the pretest measures were nonsignificant (see Appendix I), and the participants were randomly assigned to the classes which later became control and test groups, it was fair to assume that the groups

<sup>7</sup> Levene's T-tests showed homogeneity of variance assumptions were met for the production of academic formulaic sequences in an uncontrolled situation, topic-induced formulaic sequences in a controlled situation and writing ability (see Appendix I). The assumption was not met for the production of the academic formulaic sequences in a controlled situation, topic-induced formulaic sequences in an uncontrolled situation, and writing ability (see Table 3).

were equal prior to treatment. The significance and implications of these findings will be discussed in Chapter 5.

### **Qualitative Analyses from the Follow-up Interviews**

To provide a more in-depth exploration of the approaches students took at production of the target formulaic sequences in tasks, additional data were collected via follow-up interviews with some of the participants. It seemed important that a set of the same questions be asked to all the participants but that they also be asked specific questions related to their performance on the tasks used in the present study. In an effort to gain a range of students' perspectives, there were six students invited for interviews, three from the high achieving and three from the low achieving group. The high achieving students were considered those with scores in the higher band of scores and with large gains whereas the low achieving students were those that consistently performed at a lower level on all the tasks. The determining factor on the decision was the students' performance on the AFL controlled test. There were five students who agreed to the interviews - three low achieving and two high achieving. The interview focused on finding out about the students' backgrounds, in terms of academic pursuits and English language training, and, more importantly, the strategies students applied to learn and to produce the target formulaic sequences in their essays.

Al was a male student from Japan who had been in the United States for 2 months. His plan was to pursue a degree in teaching English as a foreign language in his home country. He was attending ESL classes in the U.S. to better prepare himself for the study at a Japanese university and also to improve his chances of finding employment in



his home country upon the completion of his degree. The classes at the English language institute constitute his first experience in studying English outside Japan, and more importantly, his first experience with writing extensively in English. Because he performed well on the tests measuring the production of the formulaic sequences in a controlled situation, the researcher was particularly eager to learn about the strategies he applied in producing the formulaic sequences and also the strategies for learning the formulaic sequences.

Al reported that in his home country he had experience with studying formulaic sequences (e.g., *refer to*, *according to*). When working on the C-test pretest on the production of academic formulaic sequences, he did not know many formulaic sequences; however, the language of the target formulaic sequences was familiar to him because of the reading he had done in his home country. Consequently, he tried to make educated guesses. At the end of the semester, production of the academic formulaic sequences on the C-test went smoothly and quickly because the target forms were taught in class. He felt the same about completing the C-test on the topic-induced formulaic sequences.

He reported that when he wrote the in-class essay at the beginning of the semester, his focus was on expressing his ideas on the topic; however, when writing in-class on the same topic at the end of the semester and after explicit instruction of the formulaic sequences, he focused more on the vocabulary, paying attention not only on what to say but also how to say it. He commented that he was trying to use the specific words to express his ideas. He pointed out that he participated in the in-class activities and said that they provided him with substantial practice with production of the formulaic

sequences. However, he consistently studied the topic-induced and academic formulaic sequences at home. The approach he applied to studying the academic formulaic sequences was similar to the approach he used in studying the topic-induced formulaic sequences.

Al studied the formulaic sequences through vocabulary-focused reading practice that consisted of re-reading the course materials, highlighting the formulaic sequences in the texts, and examining their form, use, and function in the texts. For the study of academic formulaic sequences, he applied this approach to reading a range of texts, specifically the newspaper and magazines. In fact, he asserted that he started noticing the phrases taught in class “everywhere” and was very surprised at the frequency the academic formulaic sequences in the texts outside the classroom context. The experience made him realize how important it was to learn these phrases to help him understand the text and also how important it was for him to use these phrases in his own writing to be clear and precise in his expression. For this reason and to further explore the academic formulaic sequences, he read the newspapers and magazines, in particular *The New York Times*, and paid close attention to reasons why the author used the constructions and the grammar of the phrases.

Al said that the class changed the way he approached reading and that because of the class he started paying attention to the vocabulary the author used, both related to the topic and those frequent in academic discourse. He said that he managed to use the academic formulaic sequences in the in-class timed essay because these formulaic sequences were “innate.” When asked about the second draft of the essay on international adoption, which used fewer formulaic sequences than the essay on genetic engineering

and was the first multidraft essay, he reported that his motivation was an issue and that given that the class instruction ended before the students were asked to submit their final drafts of the third essay, he felt he was done.

In terms of the usefulness of the in-class activities, for the purposes of using the academic formulaic sequences in writing, Al expressed a clear preference for a categorizing activity, and to the researcher's surprise was able to recall the terminology used for several categories of formulaic expressions (i.e., referential expression and obligation). It was also interesting that the learner was able to recall when he learned certain formulaic sequences: He reported that "*due to*" and "*according to*" he learned in Japan; "despite the fact" through listening to the science podcasts; and the construction "*the NOUN of*" in the writing class. It was in the writing class that he started paying attention to the phrase "*according to the*" and the fact that it was frequently followed by a common noun preceded by the definite article. In sum, Al was a language learner who was very self-aware and capitalized on the in-class opportunities to practice the use of the formulaic sequences on his own time and to explore further their use in texts. He read newspapers looking for formulaic sequences explicitly taught in class.

Another participant who was considered a high-achieving learner based on gains and performance on the C-test of academic formulaic sequences and the topic-induced formulaic sequences in the essay was Jumi, a female Japanese student planning on pursuing her degree in sports medicine at a university in the United States. She had been in the U.S. for 10 months during which time she had completed four sessions of the ESL classes at the English Language Institute. She had taken four writing classes, found writing difficult, and wanted to make a considerable improvement in the area.

Jumi reported that she studied some formulaic sequences in her native country and some multiword transitions in the previous writing classes, but many of the target formulaic sequences on the test on the production of academic formulaic sequences in a controlled situation were unknown to her. When she was asked to complete the C-test at the beginning of the semester, she was not very confident and was making guesses; however her confidence increased by the end of the semester because she was familiar with the formulaic sequences and the task. She tried to study the academic formulaic sequences through handouts given in class and also tried to use them in her essays and homework assignments. The approach she applied at studying the academic formulaic sequences was memorization. When she was writing the essay on genetic engineering, she made a conscious attempt to make her essay academic; however, when writing the essay on adoptions, the semester was nearing an end, and her enthusiasm was winding down. These factors affected her motivation in constructing an academic essay.

Jumi reported that she did not study enough prior to the C-test on topic-induced formulaic sequences. Also, she did not use the topic-induced formulaic sequences in the in-class pretest essay because she did not know much about the topic or the vocabulary to use to discuss the topic. However, she used the topic-induced formulaic sequences in the in-class posttest essay because she had studied them in class, and she realized that she could have expressed her ideas clearly and precisely had she used the formulaic sequences taught in class.

Filling in the blank and matching activities were most useful to Jumi in learning the formulaic sequences of academic prose. She found all activities useful for her writing; however, she said she did not enjoy the speaking element in the 2/1/30 activity or the

writing practice that did not have an immediate connection to her writing. For example, when discussing the pros and cons of international adoption, she was in favor of inter-country adoptions, which is why she did find useful writing an argument against foreign adoptions, a segment of the “Build an Argument” activity.

Jihan, a participant from the low achieving group, was a male student from Turkey who had obtained master’s degrees in business and engineering in his home country. He had been in the United States for almost nine months, completed four sessions of ESL classes. He had been taking English classes to improve his chances of finding employment in a prestigious foreign firm in his home country. This was the fourth writing class he had taken, and he reported that his job did not require academic writing. He felt overwhelmed with the workload in the courses for his level and was not able to focus on working on improving his essay writing as much as he had hoped. The strategy in learning vocabulary he used prior to the writing class was memorization.

When Jihan was working on the pretest on the production of academic formulaic sequences in a controlled situation at the beginning of the term, he reported that he was simply guessing. By the end of the term, he felt he had learned some formulaic sequences from the in-class activities. He said that academic vocabulary was important, but not important to him. He felt that his focus was on gaining survival skills in English not learning phrases used in academic prose. His focus was on communicating meaning not necessarily on sounding advanced in his expression. He felt that he was not ready to learn the phrases. He continued to point out that the phrases were useful and that the activities completed in class were useful, but that he needed different words not those explicitly taught in class. He decided that the vocabulary the teacher selected was not the

vocabulary he felt he needed to learn. In his writing, he used the formulaic sequences that could help him connect his ideas, some of which were those that he had already learned and those that were transferred from his native language. He also said that one of the academic formulaic sequences he used in his essay (e.g., “does not have”) was there not because he was trying to emulate academic prose but to create a grammatical sentence.

Before the C-test assessing the knowledge of topic-induced formulaic sequences was administered, Jihan focused on reading the course materials. The vocabulary he explored and focused on, was the vocabulary he self-selected either because he thought the sequences were new to him or because he found them interesting. This vocabulary was different from the vocabulary the teacher focused on. He also explained that he was interested in the topic of genetic engineering and had time to work on this paper; whereas, when he was writing the essay on the topic of international adoptions, he had family matters to attend to and the instruction had already ended. His motivation to work on the essay decreased. He explained that when he was writing essays for the class, he was focusing on creating a well-organized, unified, and coherent essay; it was a problem for him to focus on vocabulary. His approach was to think in his native language and then translate to everything to English, paying special attention to following the writing conventions that he was taught in his writing class.

Jihan found all of the activities done in class useful, but he believed that the vocabulary explicitly taught in class was not the vocabulary he needed. The in-class activity he found useful for learning how to use the topic-induced formulaic sequences in his writing was the “Build an Argument” activity. He did not find the filling-in and matching activities useful in terms of learning how to use the target forms in his writing.

However, the filling-in activity was useful when studying the academic formulaic sequences. The activities he did not find useful were dictogloss and categorizing. He explained that both activities were very difficult to complete. Dictogloss was hard because he was not able to divide his attention among the tasks--taking notes, focusing on content, and focusing on academic formulaic sequences. Categorizing was difficult because the language used to describe the phrases was difficult for him to understand. It was also difficult for him to figure out the functions of the academic phrases in the readings. He reported that there were two activities that were most useful. The first was a poster presentation, an activity the teacher herself designed to build background knowledge on the topic prior before they are asked to write. The activity was not one of the activities specifically designed for building the knowledge of formulaic sequences. The other activity was "Build an Argument." Both of these activities were rich in content and very concrete.

Another participant from the low achieving group was Jack, a male student from the United Arab Emirates. He had been in the United States for a year and three months. His goal was to continue his academic studies in the United States. Jack reported that he was familiar with most of the academic formulaic sequences because he had encountered them in books and materials he had read in the past. He added that the class helped him recall the phrases he had seen before. He repeatedly stated that the test of the production of academic formulaic sequences was an easy task to complete because it gave the students who did not study at home an opportunity to learn them learn from the test as filling in the missing parts of the words would help them remember the target formulaic sequences. He added that the activity did not require of students to "think a lot about the

answers.” He reported that prior to taking the C-test on topic-induced formulaic sequences, he read the readings on the topic and consulted on-line resources to learn more information on the topic. The strategy he used in completing the C-tests on the topic-induced formulaic sequences was the same as the one that he used in filling in the C-test on the academic formulaic sequences – guessing.

Jack reported that the class helped remember the phrases he studied before. He attended every class and applied the guessing strategy, basing it on the elements of the phrase that were provided. The time he remembered that he used the academic phrases in his writing was when the teacher asked the students to review their essays to use the academic phrases on the list students created in a sorting activity. He said that he did not use the topic-induced formulaic sequence on the topic of adoptions because he wanted to talk about adoption in general not necessarily about international adoptions.

When asked about the activities that helped him learn the target formulaic sequences, Jack replied that the activities were the poster presentations, readings, and in-class movies. What helped him the most was reading the articles in and outside of the class and watching movie clips on the topics discussed in class because the readings and the movies utilized the vocabulary he needed to discuss the topics in his own writing.

The final participant was Ju, a female student from Brazil who had a bachelor’s degree in business. She had been in the United States for 13 months. During this time, she had attended the language school for five terms moving from Level 3 to Level 7. This was the fifth writing class she had taken at the language institute. She reported that she liked writing but found it difficult to follow the conventions of academic writing. When asked about how familiar she was with the academic formulaic sequences and the topic-



induced formulaic sequences, she replied that she learned some of the formulaic sequences, such as “*in contrast to*” and “*according to*”, in the previous writing classes. When she took the pretest on the production of academic formulaic at the beginning of the semester, guessing was the main strategy she applied. On the final test, she did not guess as much as she did the first time. She was also reading the context to figure out the target formulaic sequences. She pointed out that the in-class practice helped her learn the formulaic sequences and how to complete the task. Production of the academic formulaic sequences was more difficult for her than the production of the topic-induced. She said that she had to study former on her own without being able to explain how she actually studied them. She was also not able to recall some of the in-class activities such as dictogloss that were used in teaching of academic formulaic sequences, even though she was present when the activities were done.

She said that the phrases on the topic were useful to her because they were related to the topics of the essays she wrote. She reported that the teacher used them, that she saw them used in class discussions, and that she saw them in the course readings. She asserted, “I used them in the essay because I had to use them.”

When asked about her use of the academic formulaic sequences in her essays she responded that her focus was on writing the essay and the words that discussed the content. She did not pay much attention to the academic words because she was more concerned with getting the information across not about sounding “academic.” She commented that it would have been much more helpful to her as a writer, had the teacher given the students a list of academic formulaic sequences as a reference and required them to use the phrases in their writing. Keeping a list of selected phrases that help her in

organizing the ideas of her writing was one of the strategies she employed while writing essays in English.

Ju was not confident about using the phrases. In the essay on genetic engineering, she used four academic phrases, but only two of the four were identified by the researcher prior to the interview, so Ju was asked to explain her use of the two formulaic phrases: “*on the other hand*” and “*it is necessary.*” The former she learned in a previous writing class and the latter was a direct translation from her native language. In fact, she was not aware that the phrase was one of the academic phrases or one of the phrases examined in the class. Ju used two academic phrases “*according to the*” and “*due to the*” in the essay on international adoptions, for which she reported to have learned before which seems highly unlikely given that the writing and/or grammar textbooks generally present the two formulaic sequences without the definite article.

Ju reported that she learned the topic-induced formulaic sequences on the topic of adoption in class and wrote them down from memory when asked to write the in-class essay the end of the semester. By the end of the semester she realized that it was important to use the topic induced formulaic sequences and that there were different ways to express the same notion. She explained that she used different vocabulary when she wrote on international adoptions on the pretest and on the posttest and that she changed her position. On the posttest, she made a conscious attempt to use a variety of expressions on the topic in her writing.

Ju reported that among the activities used in teaching topic-induced formulaic sequences the activity “build an argument” was the most useful. She could not think of any activities that were not helpful to her but added that the activities that were not

interesting were not useful. She felt that she would have learned more had the academic formulaic sequences had more content. Overall her experience in the class in terms of vocabulary learning was good.

## **CHAPTER 5**

### **DISCUSSION**

The chapter provides a summary and interpretation of the main findings and addresses the research questions formulated for the study. It is organized in three major sections. The first section opens with a brief introduction of the general aim of the study and a summary of the results. Further it is divided into three major parts addressing each of the three areas of focus for the present study: an investigation of the effects of explicit instruction of the academic formulaic sequences on L2 writing, an examination of the effectiveness of explicit instruction of the topic-induced formulaic sequences on L2 writing, and an investigation of the influence of explicit instruction of the formulaic sequences on the quality of L2 writing measured by the judgments of trained writing raters. The second section interprets the results of the qualitative data analysis, and the third considers the study implications for the field of L2 writing pedagogy.

#### **Discussion of the Results of Quantitative Data Analysis**

This study provides a broad picture of the effects of explicit instruction of formulaic sequences, specifically academic and topic-induced formulaic sequences, on L2 writing. In many ways, the present study confirms the findings of the previous

research and, more importantly, provides additional insights into how the explicit approach to teaching formulaic sequences affects the L2 writing.

The descriptive statistical results showed that both control and treatment groups improved over the course of the semester on every measure (i.e., production of the academic formulaic sequences in a controlled situation, topic-induced formulaic sequences in an uncontrolled situation, and essays of better quality) but one (i.e., production of academic formulaic sequences in an uncontrolled situation). Also it is not certain what the students' gains in the production of topic-induced formulaic sequences in a controlled situation because no pretests were administered. The findings confirm that some learning of formulaic sequences in a writing class does take place when the students are acquainted and reacquainted with the target formulaic sequences in readings and class discussions. The findings are particularly important because they highlight the value of providing L2 writers with opportunities to encounter the formulaic sequences through readings and class discussions.

Nevertheless, when a MANOVA was conducted to explore possible performance disparities between the experimental and control groups on their performance on the production of the target formulaic sequences in controlled and uncontrolled situations as well as on the quality of students' writing, it was found that there were overall large differences between the two groups with the students in the experimental group outperforming the students in the control group. In fact, over 50% of the variance in the students' academic performance can be attributed to the between-the-subjects variable, which was the group membership. A more complete picture of the effects of the explicit

instruction is provided by the results of the Univariate Tests that are further discussed below.

### **The Effects of Explicit Instruction of the Academic Formulaic Sequences on L2 Writing**

**Explicit instruction of the academic formulaic sequences (controlled situations).** The first research question was directed towards the investigation of possible differences in abilities to produce academic formulaic sequences in a controlled situation between the students who received (the treatment group) and those who did not receive explicit instruction (the control group) on the selected academic formulaic sequences. To examine the production of the formulaic sequences in a controlled situation a C-test was used. The C-test taps declarative knowledge and provides indirect assessment on the nature of the processes involved in language production.

The results suggest that explicit instruction of academic formulaic sequences has a positive effect on the learners' abilities to produce the target formulaic sequences in a controlled situation. The findings are in line with the results of the study by Jones and Haywood (2004) reporting an improvement in the students' abilities to produce the academic formulaic sequences in a controlled situation. The results of the present study also corroborate the findings of the study by Schmitt et al. (2004), which reports that the learners in an English for Academic Purposes (EAP) course improved their productive knowledge of the formulaic sequences over the period of 2 to 3 months of instruction, but these gains could not be attributed to the type of instruction because no control group was employed. An important contribution of the finding of the present study that employed a

control group and was conducted over a period of eight weeks is that it indicates that the explicit instruction facilitates the learning of academic formulaic sequences.

**Explicit instruction of the academic formulaic sequences (uncontrolled situations).** The present study took a step further by examining whether or not there was a significant difference in students' abilities to produce, in an uncontrolled situation (i.e., an essay), formulaic sequences frequent in academic prose (i.e., selected AFL [Simpson-Vlach & Ellis, 2010]) between the students who received explicit instruction and those who did not. The results of Univariate Tests revealed that there were no statistically significant differences between the groups on this measure indicating that the performance of the students in the experimental group did not significantly change due to the treatment. This finding matches, to a certain extent, the results of the research conducted by Cortes (2006) on L1 writers, as well as Jones and Haywood (2004) on L2 writers. The students in Cortes' (2006) study (who were native speakers of English enrolled in a university-level history course and received explicit instruction over a period of 10 weeks on a set of lexical bundles<sup>8</sup> most frequent in their discipline) did not show progress in frequency or mastery of use of the target lexical bundles in their written assignments. The researcher inferred, for the most part, that the students possibly needed more instruction than that provided in the five minilessons in the semester and that they needed to engage in activities that provided more exposure to the ways in which the formulaic sequences are used in context.

Similar to the findings of Cortes (2006), the students in the Jones and Haywood (2004) study showed a lack of progress in the production of the formulaic sequences

---

<sup>8</sup> Lexical bundles are defined as combinations of three or more words that most frequently reoccur in a register (Biber et al. 1999, p.990).

between Essays 1 and 3. The researchers attributed the lack of progress to two factors. The first factor was the short period of instruction (i.e., two weeks) on a type of genre for the second essay; the second factor was the limited support the students received in the process of writing of the final essay. The students were supplied with more reading materials and engaged in more in-class discussions in preparation for writing the first essay.

The present study attempted to control for the aforementioned factors in the following ways: The students explored the argumentation as a writing genre for the entire instructional period of 8 weeks; they received multiple readings in preparation for writing of the first and the final essays; and they engaged in weekly activities that provided extended exposure to the target formulaic sequences. Yet, the present study corroborates the findings reported by the previous research (Cortes, 2006; Johns & Haywood, 2004). Considering the fact that the present study attempted to correct for the reported limitations of the previous research and still obtained similar results, seems to suggest other factors may have contributed to the lack of the effect of the treatment on the students in the experimental group.

While it may be the case that the lack of observed effects of the explicit instruction on L2 writing for the students in the treatment group may be that the students need more exposure to and more practice with the use of the target formulaic sequences as noted by Jones & Haywood (2004) and also by Cortes (2006) in the L1 context, it is also possible that with the limited time available for explicit instruction, the students need to be engaged in a specific type of practice in order for the treatment to be realized and that other factors such as motivation may be associated with the observed students'



performances. This hypothesis is set forth based on the information gathered in the follow-up interviews with the students from the treatment group.

In the interviews, all of the students stated that the extensive practice with cloze-type activities (e.g., matching and C-test) was useful, but for the majority of the informants this type of practice did not seem sufficient. They could not bring their knowledge of the academic formulaic sequences into active use in their L2 writing. In fact, four out of five students reported that more activities that aim at the production of the target academic formulaic sequences in meaningful contexts and particularly those with a direct application to the students' texts would have been useful to their writing practice. Some possible ways on how to go about creating opportunities for learning of academic formulaic sequences that may be more useful to student writing are discussed in the final section of the present chapter entitled "Implications for L2 Writing Pedagogy."

What the role of motivation in the development of formulaic language has not been yet fully understood. In fact, a longitudinal study conducted by Schmitt et al. (2004) with a group of academically-bound ESL students reported no effect for a motivation variable on the production of formulaic sequences. In light of the findings and what is known about the effects of individual differences on the language learning from Dörnyei (2002) and Dörnyei and Csizér (2002), the researchers conclude that motivation, among other individual factors, may affect the learning and production of the formulaic sequences formulaic through interplay with other factors. The prediction of Schmitt et al. is supported by the findings of a case study by Dörnyei, Durow, and Zahan (2004). The researchers explored the success of formulaic sequences acquisition of four successful

and three unsuccessful learners who were post-graduate students at an English-speaking university via quantitative (i.e., aptitude and motivational tests) and qualitative measures (i.e., interviews). The findings from the interviews indicate that motivation in interaction with other factors (i.e., language aptitude and sociocultural adaptation) may contribute to the production of formulaic sequences. The results from the interviews from the present study seem to provide further support on the possible indirect influence of motivation on the production of formulaic language, in particular the written production of the academic language.

Indeed, the results indicate that it is possible that the students' production of the academic formulaic sequences was influenced by the lack of motivation to craft an academic argumentative essay. Each of the five students who was interviewed after the treatment, either explicitly stated or implied that their motivation to give their best on the final draft of the final essay had decreased. Their reasons pointed to the fact that the final essay was to be submitted for evaluation after the last day of instruction and that in their minds the term had already ended. The students felt that they already had an idea what their final grades were going to be and that their failure to "sound academic" on the final essay would not necessarily affect their grade in the course in a negative way. Even the high achieving students reported that their motivation to focus on academic vocabulary was lacking on the final draft of Essay 3.

The discussion of results now turns to possible explanations for the differences in the performance between the groups in the present study that stood out in the descriptive statistics report. The results of descriptive statistics revealed that a) the students performance on the production of the academic formulaic sequences decreased for both

groups from Essay 1 to Essay 3 as depicted in the group means (i.e.,  $M = 14.12$  to  $M = 11.18$  for the control and  $M = 17.30$  to  $M = 8.77$  for the experimental); b) the performance of the experimental group decreased more than the performance of the control group; and c) the control group outperformed the experimental on the production of academic formulaic sequences on the Essay 3.

First, the decrease in scores noted in the present study, though a disappointment, should not come as a great surprise because Jones and Haywood (2004) noted a similar trend. In their study, five out of the six participants who submitted their first and final multidraft essays for evaluation had scores on the final essays that went down. The researchers also looked at the possible change in accuracy of the form of the formulaic sequences in students' essays by comparing the mean scores per sequence per student and found out that three out of the six students made slight improvements over the course of instruction. (The present study did not examine the mean scores per formulaic sequence, but future research should look at possible differences.) The researchers characterized the findings as inconclusive and discussed only the reasons behind the lack of evidence of progress on overall writing.

Some possible explanations for the decreased performance from Essay 1 to Essay 3 can be offered based on the information collected through the individual interviews with a group of students from the treatment group and the reevaluation of the input students received in the class. The first finding may be explained in light of the students' decreased motivation due to the conditions under which the final essays were collected. There is a good chance that both groups were more motivated to perform well on Essay 1 as it was the first major assignment for the course for which the students received a grade

that figured into the final grade for the course. The final essay (Essay 3) was submitted after the last day of instruction.

Considering that “. . . time is needed . . . , for attention to vocabulary” (Raimes, 1985) and that time was limited in the process of writing the first draft of Essay 3 (it was an impromptu in-class essay produced under time constraints), it is possible that the first draft for the Essay 3 was composed with a limited focus on vocabulary, including academic vocabulary, specifically the academic formulaic sequences. If this is the case, then there is a possibility that the scores on the final essay decreased because of the combination of two factors: 1) students’ lack of motivation to produce academic prose and 2) the initial constraints on time in drafting Essay 3, resulting in inferior use of academic vocabulary when compared to Essay 1.

Another reason for the decreased performance on the production of academic formulaic sequences on Essay 3 may be due to the frequency of the academic formulaic sequences in the reading materials on the topics. There were 58 target formulaic sequences used in the readings completed in preparation for writing on the topic of genetic engineering and 16 target formulaic sequences in the materials examined prior to writing on the topic of international adoptions. In light of the fact that there were almost four times more academic formulaic sequences used in one set of the reading materials than in the other, it is possible that the frequency of the academic formulaic sequences in the reading materials influenced the students’ production of the academic formulaic sequences. It does not seem that there were the topic effects on the frequency of the formulaic effects in the texts because the reading on the two topics in the textbook contained almost the same number of the academic formulaic sequences.

The findings from the present study suggest that the performance in the production of the academic formulaic sequences started higher but decreased more for the participants in the experimental group than the participants in the control group. Even though, the differences between the groups were not statistically significant, as revealed by the independent T-test and Univariate Tests, the differences seem worthy of attention. The finding that the experimental group outperformed the control on the production of academic formulaic sequences on Essay 1 may be due to the combination of factors: a) the high motivation to perform well on the first written course assignment as reported by the students in the follow-up interviews; and b) a possible increased awareness of the importance and frequency of the formulaic sequences in academic writing (Jones & Haywood, 2004) that was noted by one of the informants. The student explained that the academic formulaic sequences were “everywhere.” The large decrease in the performance of the experimental group may be due to the fact that the experimental group started higher than the control and then decreased at the rate similar to the control group.

Finally, the results of the descriptive statistics suggest that the scores for the students in the control group outperformed the students in the experimental in their ability to produce the target academic formulaic sequences on the final essay. While this finding is somewhat puzzling to the researcher (i.e., why didn't explicit instruction of formulaic sequences in controlled environments have more of an influence on the L2 writers' production of the academic formulaic sequences in a final essay?), it is possible that the scale used in scoring of the academic formulaic sequences does not consider any other issues in form of the formulaic sequences except the following: the issues with

inflectional morphology, substitution of a preposition (e.g., *in the other hand* instead of *on the other hand*); and issues with pronoun use and omission of a function word inside the phrase (e.g., *as result* instead of *as a result*). There were instances of formulaic sequences in the texts of the students in the experimental group that were excluded from consideration for two main reasons: a) because the process of evaluation of the essays would have been very difficult to carry out, as it would create an extremely large pool of phrases for which to search the essays and b) because the academic formulaic sequences in the list offered by Simpson-Vlach and Ellis (2010) are not listed in the form to allow insertions or alternatives. Some examples include *a few amount of* and *a huge amount of, the increase number of* as well as instances such as *there are two* which appear to be students attempts to produce the following academic formulaic sequences “*the amount of,*” “*a (large) number of,*” and “*there are three*” or possibly “*there is a*” respectively.

### **The Effects of Explicit Instruction of Topic-Induced Formulaic Sequences on L2 Writing**

The second area of interest of the present study was an investigation of the effects of explicit instruction of the topic-induced formulaic sequences on L2 writing. The research questions that motivated this portion of the study examined whether or not there was a significant difference in students’ abilities to produce formulaic sequences in topic discussion between the students who received explicit instruction and those who did not in a controlled situation (i.e., a C-test) and in an uncontrolled situation (i.e., an essay). The results of the Univariate Tests reveal statistically significant differences on both measures (i.e., controlled and uncontrolled situations). These findings, even though

related to the topic-induced formulaic sequences, come as an expected result to researchers who have been suggesting that a possible way to help students learn formulaic sequences, academic to be specific, would be explicit instruction (Coxhead & Byrd, 2007; Simpson-Vlach & Ellis, 2010). Having said this, it is nevertheless, important to highlight that the findings such as these were not a result of superficial or random instruction of the topic-induced formulaic sequences. They were a result of a moderately intensive, focused, and carefully planned explicit instruction of the formulaic sequences that were assumed to be useful to L2 writers because they had an immediate application to their writing.

The finding that the explicit instruction has a positive effect on the students' abilities to produce the formulaic sequences in a controlled situation, namely a C-test, probably does not come as a surprise to anyone as it is in line with the reports of the previous research in the area of learning of academic formulaic sequences (Jones & Haywood, 2004; Schmitt et al., 2004) and the findings of the present study that have been discussed in the previous section. An important contribution of this finding to the existing evidence in support of explicit instruction of formulaic sequences is that the explicit instruction can facilitate the learning of not only academic formulaic sequences but also another type of formulaic sequence, which is in the current study, topic-induced formulaic sequence.

The second research question investigated whether or not there was a significant difference in students' abilities to produce formulaic sequences in topic discussion between the students who receive explicit instruction and those who do not in an uncontrolled situation. The answer to the question is probably the most notable in the

present study because it indicates that explicit instruction of this type of formulaic sequences had a facilitative impact on the learners' production of the target formulaic sequences in their timed in-class essays. Because the students had access to the reading materials as they were writing the in-class essays, it cannot be argued strongly that the students recalled the sequences from memory; however, that the students retrieved the topic-induced formulaic sequences from memory should remain a possibility because they were working under time constraints. If it is the case that selecting and using appropriate vocabulary in writing requires time and attention (Reimes, 1985), then it just might be the case that the students were using the target formulaic sequences in free production. This finding makes an important contribution to earlier work. It offers empirical evidence to the existing literature that suggests that learners be taught explicitly the vocabulary they need (i.e., Coxhead & Byrd, 2007; Folse, 2008; Nation, 2005; Schmitt, 2000).

Although the objective of the present study was not to examine the outcomes of the explicit instruction relative to the type of a formulaic sequence (i.e., academic and topic-induced formulaic sequences), the difference in the effects of treatment between the academic and topic-induced formulaic sequences can hardly be ignored, especially in light of the fact that the instruction on the topic-induced formulaic sequences lasted for four days and the teaching of the academic extended over the period of 8 weeks. It is also risky to discuss the effects of explicit instruction on the development of formulaic knowledge for productive use in the presence of an extraneous variable such as the availability of sources during writing; however, having considered the aforementioned study findings, it may be that the formulaic sequences that are rich in content are learned



faster than those that are not. In addition, it may help explain why the students in the Jones and Haywood (2004) study learned to produce the noun phrase with the of-phrase (e.g., the presence of) first and why this particular category of formulaic sequences stood out among others for the high-achieving writer as the one present everywhere in the texts of expert writers.

### **The Effects of Explicit Instruction of the Academic and Topic-induced Formulaic Sequences on the Quality of L2 Writing**

The third section of the study addressed the question of whether or not there was a significant difference in the overall quality of students' essays as measured by the trained essay raters' judgments between the students who received explicit instruction and those who did not. The study results showed no statistically significant difference between the quality of the essays produced by the students who received and those who did not receive explicit instruction on the formulaic sequences. Given that the effects of treatment were observed on the students' performance on the production of the topic-induced formulaic sequences only, it seems that only utilizing this type of formulaic sequence does not influence the overall quality of their writing represented by a cumulative score on individual scores on five aspects of writing (i.e., content, organization, vocabulary, language use, and mechanics).

The descriptive statistics show that the control group performed slightly better than the experimental on a writing task at the beginning of the semester but that the experimental group outperformed the control group after the treatment. Also the mean

scores on the overall quality of writing of the low performing and high performing students by group reveal that the low performing writers in the experimental group made greater gains than their counterparts in the control group. The similar trend is observed in high performing writers in the experimental group, but their scores at the beginning of the semester were only slightly lower than the scores for the control group; consequently, the gains, even though larger than the gains of the high performing students in the control group, do not stand out. These findings indicate that both low and high performing writers benefited from the explicit treatment of the topic-induced and academic formulaic sequences, but that the instructional intervention may be particularly valuable to the low performing writers to produce better quality prose.

When mean scores for each of the five aspects of writing (i.e., content, organization, vocabulary, language use, and mechanics) were examined, it was revealed that that the experimental group outperformed the control on two aspects of writing: vocabulary and language use. The former is particularly important given that the descriptive statistics show that the experimental group was slightly inferior to the control in terms of their overall performance on vocabulary in writing prior to treatment. While these findings may be encouraging in terms of the effects of explicit instruction of formulaic sequences in writing, they also suggest that the slight gains in vocabulary and language use may come at a cost. The students in the experimental group may have improved their vocabulary and grammar for the purposes of writing, but seemed to lag behind the control group on the other areas of writing: content, organization, and mechanics. How costly the intervention may have been is displayed in the results on the students' performances on the organization aspect of writing: the experimental group

started higher but fell behind the control group after the instructional intervention.

### **Discussion of the Results of Qualitative Data Analysis**

This section of the chapter offers a discussion of the results obtained in the follow-up interviews with a subset of students from the treatment group, which were conducted to glean insights into the strategies ESL learners use in producing the formulaic sequences in their writing. The interview data provided very important information relevant to the interpretation of the findings for the research questions. Additionally, they provided more insights into the approaches students took to the learning of the target formulaic sequences and producing the target formulaic sequences in writing.

In the discussion a reference will be made to the high and low achieving students to refer to the students who were identified as such based on their performance on the C-test on academic formulaic sequences. Because there were multiple tasks in the study and some students maintained their performance across tasks while other students' performances varied depending on the task and task conditions, there will be a reference made to the students who received high scores or performed well and those who receives low scores or performed poorly on particular tasks. The reference to these students may cross the previously noted distinction made between the high and low achieving students.

One of the findings from the interviews suggests that students' performances on the production of the formulaic sequences on the C-tests were dependent on whether or not they invested time and effort in studying the target formulaic sequences outside the class. The students who obtained high scores on the tests concurred that studying the

formulaic sequences by reviewing the handouts and focusing on the vocabulary while rereading the texts provided by the teacher influenced their abilities to perform well on the tasks. The students, whose performances on the tests were low, reported that they did not engage or did not engage enough in a focused study of formulaic sequences outside of class.

Second finding of the follow-up interviews suggests that the students seemed to be more likely to use the explicitly taught formulaic sequences in their compositions if they perceived them immediately applicable to their own writing. The students who used the explicitly taught formulaic sequences in their essays reported that they found the target formulaic sequences useful to their writing because they helped them state their ideas clearly. The students who used the formulaic sequences less frequently asserted that they did not find the formulaic sequences useful to their own writing. With respect to the production of the academic formulaic sequences and the perceived necessity of use, several students explained that that they did not fully understand the functions and usage of the academic formulaic sequences, suggesting that the explicit instruction on the academic formulaic sequences needed improvement.

Another finding that emerged from the one-on-one interviews was that the learners' decisions about which formulaic sequences to perceive as useful to their writing and which to perceive as not useful were influenced by the students' awareness of the frequency and functions of the formulaic sequences in the reading materials. How important to learning of the formulaic sequences an awareness of their frequencies and functions is best described in the behavior of one of the high achieving students who reported that the writing class changed his approach to reading. Due to the explicit

instruction that raised his awareness of the importance of formulaic sequences, specifically the academic formulaic sequences, he started to engage into reading with the purpose to attend to the language not only the message in the text and to investigate himself how the formulaic sequences are used by expert writers. Other students reported that the reason they used the topic-induced formulaic sequences in their writing was because they were “everywhere” in their readings, in the handouts, in the movie clips they watched, in the speech production of their writing teacher.

With respect to the learner’s production of academic formulaic sequences in writing, the findings of the interviews suggest that production was affected by the students’ motivation to emulate academic writing. Most of the time, the production of academic formulaic sequences was a result of a conscious attempt on the part of the student to sound academic. The low achieving students reported they had referred to the formulaic sequences lists, either self-compiled or provided in the handouts given by the teacher, as they wrote their essays. The high achieving students, for the most part, tried to recall the academic formulaic sequences from memory. When a conscious attempt to sound academic was absent, yet the academic formulaic sequences were produced in writing, the academic sequences seemed to be a result of a direct transfer from the students’ native languages (Altenberg & Granger, 2001; De Cock, 2003; Granger, 1998; Paquot, 2008) or an attempt to produce a grammatical construction in English.

Even though the motivation to sound academic may have an effect on students’ production of the academic formulaic sequences, based on the findings from the interviews, it seems to be in itself an unstable factor. The students reported that their motivation to emulate academic writing on the final essay decreased largely because their

final drafts had to be submitted after the last day of instruction, which was, in their minds, after the course was over.

The findings of the follow-up interviews revealed that high achieving students applied direct methods to studying the target formulaic sequences, such as reading with a focus on vocabulary or memorization, and the majority of the low achieving students seemed to employ rather indirect strategies to the learning of the target formulaic sequences. To illustrate, two of the low achieving students reported that reading for content was their preferred and sufficient method to studying vocabulary including the formulaic sequences, as well as watching movie clips and preparing a presentation.

With respect to the activities used in explicit teaching of the target formulaic sequences that the learners found useful to their writing, the informants generally reported that the activities with a direct application to their essays seemed to be most useful to their writing. They asserted that had there been more of these activities included in the instruction of the academic formulaic sequences, it would have been more useful to their writing.

### **Implications for L2 Writing Pedagogy**

There are several general implications on L2 writing pedagogy that are proposed based on the overall findings of this study.

First, it seems that explicit instruction of the formulaic sequences when provided alongside opportunities that allow indirect learning of formulaic sequences, such as reading articles to springboard their writing, would benefit L2 students across the levels of their writing ability. The less skilled writers may ultimately improve the overall quality

of their writing by latching on the vocabulary to express their ideas clearly which may not be the case if the instruction is absent. The more skilled writers may begin paying more attention to the words they employ in their own writing and investigating for themselves how the sophisticated expressions are used by expert writers. For both, it may raise their awareness of the frequency and functions of the formulaic sequences in writing to, in the words of one of the informants, “change the way [students] read.”

When integrating instruction on formulaic sequences, writing teachers need to be strategic in the selection of the formulaic sequences for explicit instruction and meticulous in how they carry out the explicit instruction. With respect to the selection of formulaic sequences, it seems that the students are more likely to use in their writing those formulaic sequences that they perceive as important and necessary to their composition, those formulaic sequences that serve a particular purpose in their text. For this reason, when the teacher selects the formulaic sequences, specifically the empirically derived academic formulaic sequences, it would be useful to consider the requirements for a particular assignment in the process of selection of the target formulaic sequences. For example, if the writing assignment was an argumentative essay in which students often need to compare and contrast two sides on one issue, provide examples, discuss causation and in the end present their position, the teacher could focus on the formulaic sequences that fulfill the aforementioned functions. This is not to say that all the other formulaic sequences would not be made salient in the students essays. On the contrary, the fact that they are all identified in the test for the learner may be useful in raising students’ awareness of their frequencies of use and the purpose each serves in a text of an expert writer. In other words, instead of informing students of how frequent and functional

formulaic sequences are in a text, it would be beneficial to exemplify their actual use in prose, and in terms on which formulaic sequences to teach through a series of activities, it seems beneficial to focus on those students would need in their own writing, and possibly on those with which students may not be already very familiar to help them increase the breadth of their vocabulary knowledge. Also, it may be difficult, if necessary at all, to teach all of the academic formulaic sequences offered in a list, so the teachers should take comfort in knowing that probably the most long-term achievement the explicit instruction can provide to a writer is the raised awareness of the functions and functions of formulaic sequencers in a text.

The teachers should not lose sight of the fact that the vocabulary choices expert writers make as they write is heavily influenced by the purpose of their writing. Therefore, the students need to be taught that each of the academic formulaic sequences serves a specific purpose in a texts, that is by an expert writers selected with a specific purpose in mind, and that based on the selection of the words the students themselves have chosen, they may or may not be conveying their intended message. It seems that the teachers would want to consider holding writing conferences with students in which they would specifically devote some time to a review of individual student's use of the formulaic sequences. Coxhead and Byrd (2007) seem to prefer to use the term "feedback" in their discussion of how the teachers should go about teaching vocabulary of academic prose, but in this document, the researcher chooses to focus on the feedback student receives in individual writing conferences with their teacher. If the use of the academic formulaic sequences is so closely associated with the students' perceived need for their use, then it would be premature to conclude that all the students in the class would have



the same needs. While the teacher may choose to target the academic formulaic sequences that serve the purposes of many in the class, as suggested previously, it seems necessary to work with students one on one, examining their essays to praise students' target-like use of formulaic sequences to encourage and to discuss the missed opportunities for their use and possible misuse to raise their awareness of the purpose each of the formulaic sequences serves in a text. This approach may appear very demanding on teacher's time, and highly likely on their training; however, if writing instruction at an institution is carefully planned to prepare students to focus for the demands of academic writing, this approach may be necessary.

In addition, with respect to academic formulaic sequences, it appears that the students may learn faster the formulaic sequences to which meaning can be associated. For this reason, it may be useful to focus on teaching explicitly the referential expressions or the formulaic sequences in the form of "the noun of" first.

Relative to the approach to explicit teaching, teachers would do well if they contextualize and plan carefully explicit instruction to align it with noticing, retrieval, and generation, the three principles necessary for vocabulary learning (Nation, 2001). In addition, in the approach, they would need to incorporate writing activities that focus on building the knowledge of formulaic sequences and are as closely as possible related to the immediate needs of the students and their class assignment. In so doing, they may help students increase their awareness of the necessity of use of the target formulaic sequence and provide the learners with a more meaningful context for the use of the target forms.

There may be the case that the materials students read in preparation for writing focus rather on building their background knowledge than on their academic language. However, such a situation, though discouraged, may be amended by providing students with models that utilize the academic formulaic sequences they are expected to use in their prose. This can be achieved by refereeing to the corpus, for example COCA (Davies, 2008-) as a resource for models or the references for the models. For each query, a formulaic sequence in the present case, COCA offers two contexts of use: limited and expanded. The limited context presents the formulaic sequence in one line of text; but the expanded context offers usually five to six lines of the text surrounding the formulaic sequence. The extended context may offer enough contexts for an examination of the function of a formulaic sequence. In case that it does not, the teacher may search the corpus for the model that illustrates the function of the formulaic sequence in the expanded context, or the teacher may obtain the reference information and look for the model in the library.

In addition, there is often not a strong correlation between which formulaic sequences teachers (and vocabulary researchers) perceive as necessary to L2 students' writing and which formulaic sequences L2 students themselves perceive as needed in their compositions. Based on the findings of the study, the teachers may align their students' views with their own if the explicit instruction on the formulaic sequences that the teachers find necessary is provided in meaningful contexts that relate to students' writing.

Above all, if it is not the writing class that the students are given specific direction as to which formulaic sequences are those they are expected to use and the time given to

students to figure this out on their own via reading is limited, one may ask where it is that the students will have an opportunity to learn and learn about the language that they themselves are expected to produce.

## **CHAPTER 6**

### **CONCLUSION**

The final section offers concluding remarks, discusses the limitations of the study accompanied with directions for further research and implications for further research and finally, considers the study implications for the field of L2 writing pedagogy.

#### **Concluding Remarks**

Formulaic language is becoming an increasingly important topic in applied linguistics. In fact, it is one of the fastest growing areas of research in applied linguistics (Ellis, 2009). Such an interest is often accompanied by many questions some of which concern the factors that facilitate the students' production of formulaic sequences in writing.

The present study was designed to answer the broad question of what effect explicit teaching of formulaic sequences has on L2 writing. The questions was explored through an examination of the effects of certain treatments on ESL students' abilities to produce formulaic sequences, in particular, academic and topic-induced formulaic sequences, in controlled (i.e., C-test) and uncontrolled (i.e., essay) situations, as related to their abilities to produce better quality writing.

The study found that the students were able to produce academic and topic-induced formulaic sequences in controlled situations, namely, a C-test. In other words, their declarative knowledge of the target formulaic sequences was improved as a result of explicit instruction.

In addition, the study found that the explicit teaching did affect the production of the academic formulaic sequences in an uncontrolled situation, namely, an essay. Most importantly, the study found that students learned to produce the topic-induced formulaic sequences in the writing, but this finding needs to be treated with caution. Because the students had the reading materials that contained the topic-induced formulaic sequences at hand, it could not be claimed that the students recalled the target sequences from memory. What we do know is that students were familiar enough with the formulaic sequences to recognize their usefulness and employ them in their own writing. As expert writers use different sources as springboards for their own writing, the skills of recognizing and using formulaic sequences from sources, even if not from memory, should not be discounted. Finally, although the study found no statistically significant effects of explicit instruction on the students' ability to produce compositions of better quality, there were indications that both high and low performing students benefit from explicit instruction and that the treatment may be particularly useful to the low performing students to improve their writing.

The findings from the interviews conducted with a selected group of students from the treatment group indicate that the students' abilities to produce the academic and topic-induced formulaic sequences in their writing may be dependant of the students' perceived need to use the target formulaic sequences in their compositions. With respect

to the students' abilities to produce the academic formulaic sequences in writing, the findings indicate a relationship between the students' awareness of the frequency and functions of formulaic sequences in academic prose and the students' motivation to sound academic in their writing. If the latter is absent, the students may be less likely to produce the academic formulaic sequences in their writing.

### **Study Limitations and Implications for Future Research**

Although every attempt was made to eliminate design and analytical flaws, there were inevitably some limitations, and these should be taken into account when further research is designed.

The goal of the present study was to investigate the effects of the explicit teaching of formulaic sequences on L2 writing. The study was designed as a quasi-experimental research study that employed an experimental group that received the treatment and a control group that did not. While the study attempted to set up favorable conditions for explicit vocabulary learning for the experimental group (see p. 62), the objective for the control group was to maintain the conditions for vocabulary acquisition that are generally provided in a traditionally taught writing class (i.e., readings and class discussions). The results of the present study indicate that the students' in the control group (i.e., group exposed to the target formulaic sequences through the readings and class discussions) did improve their production of the target formulaic sequences over the course of the semester. In light of these results, it would be very useful to conduct an additional research study that contrasted the effects of an explicit instruction approach to the teaching of formulaic sequences, similar to the one in the present study, to an approach

that took into account the conditions that are known to facilitate incidental vocabulary learning as reported in the literature on vocabulary acquisition (e.g., Nation, 2001; 2005; Woorland, 2000). In this case, further study would need to ensure that the exposure group was provided with opportunities that allowed for multiple encounters with the target formulaic sequences in meaningful contexts (Nation, 2001; 2005). Specifically, the control group would need to be exposed to the reading materials in which the exposure to each formulaic sequence was at a minimum of seven to attain the levels of receptive vocabulary acquisition (Woorland, 2000). There would also need to be some type of accountability in place to assure reading of the selected texts. In addition, it would be important to take into account the frequency of repetitions of the target formulaic sequences in in-class interactions with the teacher. Such a study would provide a better picture of the differences in the approach to teaching formulaic sequences and be a useful addition to the current body of literature on the topic.

Second, the study followed a group of participants at an intermediate level of English proficiency in an Intensive English Program who had various goals for the future. Many of the students confirmed that they wanted to continue their academic education in a country where English is considered an official language. However, there were some students in the study whose future plans did not include academic pursuits. In fact, two of the students whose goals were to look for employment after they completed their ESL courses fell in the group of the low achieving students. These students reported difficulties with motivation in learning the academic formulaic sequences. For this reason, the instruction of the academic formulaic language may have been better suited for the students enrolled in university ESL writing courses whose academic aspirations

have been defined. Future studies could investigate whether explicit instruction that is contextualized and aligned with students' needs has a facilitative effect on the written language of the students' who have already stepped into an academic setting.

Third, in an effort to mimic the conditions in which students most frequently write with the sources within arms' reach, the study design prevented from gathering the information on the effects of explicit instruction on the students' abilities to target formulaic sequences in free production. In addition, the students' previous knowledge of the target formulaic sequences was not taken into account. An answer to the question whether or not explicit instruction of new formulaic sequences that the students find useful have an effect on their ability to use them at will would fill a gap in second language vocabulary acquisition and pedagogy.

Fifth, because of the lack of empirically-based methodology for the instruction of formulaic sequences (Granger, 1998; Jones & Haywood, 2004) and the reliance on the findings of previous empirical research on the benefits of multiple encounters with or without retrievals of the target vocabulary (Folse, 2008) the approach to instruction of academic formulaic sequences focused heavily on the production of the target formulaic sequences in controlled and less on the production of the target formulaic sequences in uncontrolled situations. The findings of the present study indicated that the activities on production of the academic formulaic sequences in meaningful contexts and those that have an immediate application to the students' writing would have been useful to their writing. As potential future studies seek to explore the effects of explicit instruction of academic formulaic sequences, they could plan the explicit instruction so that it balances



better the practice in production in cloze-type activities and tasks that apply to students' writing.

Sixth, it may be worth to reexamine the presentation of the academic formulaic sequences offered in Simpson-Vlach and Ellis (2010) in order to account for the possible alternative forms such as insertion of modifiers within the sequence such as "*a (modifier) number of*" instead of "*a number of*" and allow for an alternative word of the same word category such as "*there is/are (number)*" instead of "*there are three.*" Such an approach would possibly validate the use of a less conservative rating scale for evaluation of the production of the academic formulaic sequences than the one used in the present study.

Finally, one very important piece of data has been collected at, what seemed to be, the most unfortunate time in the term: The students' Essay 3 was collected after the final day of instruction. This was the time when the students' motivation to perform well in class decreased because they felt the instruction was over and their grades, for the most part, had been demined. It would be useful to examine further the effect of motivation to "sound academic" on the students' production of the academic formulaic sequences in student academic writing. Such research would inform the L2 pedagogy of how better to set up the optimal conditions for students to learn to use and have an opportunity to practice to use in their compositions.

## APPENDIX A

### CORE ACADEMIC FORMULAS AFL FOUND IN THE READING MATERIALS OF AN ADVANCED WRITING CLASS ARRANGED BY PRAGMATIC FUNCTION

#### Group A. Referential Expression

##### (1) Specification of attributes.

##### *(a) Intangible framing attributes*

in response to

in terms of

is based on the

point of view

the ability to

the fact that

the fact that the

the form of

the issue of

the meaning of

the presence of

the problem of

the process of

the role of

the use of

***(b) Tangible framing attributes***

as part of a

part of a

part of the

parts of the

the amount of

the area of

the level of

the part of

the rate of

value of the

***(c) Quantity specification***

a number of

a series of

a set of

of these two

the number of

there are three

**(2) Identification and focus.**

as an example

in this case

is for the

is not a/an

is not the

is that the

is that there

it can be

it does not

it is not

is to be

referred to as

such as the

that in a

that there are

that there is

there is a

there is no

this is not

this type of

this would be

**(3) Contrast and comparison.**

as opposed to

have the same

of the same

the same as

**(4) Vagueness markers.**

and so on

**Group B. Stance Expressions**

**(1) Hedges.**

are likely to

it may be

likely to be

may not be

more likely to be

**(2) Epistemic stance.**

according to the

**(3) Expression of ability and possibility**

can be used

can be used to

to use the

**Group C. Discourse organizing functions**

**(1) Topic introduction and focus.**

what are the

whether or not

**(2) Topic elaboration: cause and effect.**

as a result

as a result of

because it is

due to the

in order to

so that the

the effects of

the result of

whether or not

**(3) Discourse makers.**

and in the

as well as

at the same time

in other words

## APPENDIX B

### WRITTEN ACADEMIC FORMULAS AFL FOUND IN THE READING MATERIALS OF AN ADVANCED WRITING CLASS ARRANGED BY PRAGMATIC FUNCTION

#### Group A. Referential Expression

##### (1) Specification of attributes.

###### *(a) Intangible framing attributes*

depend on the

depending on the

depends on the

###### *(b) Tangible framing attributes*

###### *(c) Quantity specification*

a large number of

in a number of

in most cases

there are no

there are several

two types of

##### (2) Identification and focus.

does not have

has also been

his or her

it has been

none of these

there has been

they did not

they do not

which can be

**(3) Contrast and comparison.**

on the other hand

the other hand

**(4) Deictics and locatives.**

at this stage

the United Kingdom

**(5) Vagueness markers.**

**Group B. Stance Expressions**

**(1) Hedges.**

it is likely that

appear to be

are likely to

as a whole

less likely to



**(2) Epistemic stance.**

be argued that  
 been shown to  
 if they are  
 take into account

**(3) Obligation and directive.**

needs to be  
 to ensure that

**(4) Expression of ability and possibility.**

are able to  
 be used to  
 can also be  
 their ability to  
 to carry out

**(5) Evaluation.**

important role in  
 it is difficult  
 it is important  
 it is impossible  
 it is necessary  
 the most important

**Group C. Discourse organizing functions****(1) Metadiscourse and textual reference.**

the next section

**(2) Topic elaboration: cause and effect.**

for this reason

## APPENDIX C

### TOPIC-INDUCED FORMULAIC SEQUENCES BY ALPHA LISTING BY THEME

Topic	Formulaic Sequence
Grey wolf	A natural/big predator
	Allow farmers to protect their livelihood/livestock/personal property
	Associate wolves with wilderness
	Compensate ranchers/farmers for livestock/losses
	Defenders of wildlife
	Endangered/threatened species
	Grey wolf reintroduction program(s)
	Keep the wolves away from humans
	Offer compensation to (ranchers) for livestock/losses
	Pose a danger to (humans)
	Put the wolf back on the endangered species list
	the recovery efforts of the wolf/wolves
	The slaughter of wolves/livestock
	To remove the wolf from the endangered species list

---

	Wolf attacks on dogs/humans
	Wolf habitat
International	a victim of violence/victims of violence
Adoption	adoption agencies; corruption in a country/adoption criteria for adoption criteria for adoption; foreign adoptions; inter-country adoptions; international adoptions; orphaned children orphaned children; place a child for adoption place a child in a foreign family prospective adoptive parents; reopen adoption to foreigners requirements for adoption to be adopted into; to be placed with a family/families

---

## APPENDIX D

### BACKGROUND SURVEY

**Student's name** \_\_\_\_\_

**Answer the following questions:**

1. How old are you? \_\_\_\_\_ years \_\_\_\_\_ months
2. Where are you from? \_\_\_\_\_
3. What is your native language? \_\_\_\_\_
4. How long have you been in the USA? Years \_\_\_\_\_ Months \_\_\_\_\_
5. Why are you taking English classes at the ELI? Choose one of the following answers by putting a checkmark (i.e. ✓) in the appropriate box OR fill in the blank:

- To prepare me for college/university in the United States or another English-speaking country.
- To prepare me for a college/university in my home country.
- To help me find employment in the United States or another English-speaking country.
- To help me find employment in my home country.
- Other: (please specify) \_\_\_\_\_

6. How do you feel about writing in English? Do you enjoy writing, find it stressful, or both?

## APPENDIX E

### THE C-TEST ON PRODUCTION OF THE ACADEMIC FORMULAIC SEQUENCES

Student's Name \_\_\_\_\_

Date: \_\_\_\_\_

#### Fill- in exercise

**Directions:** Read the extracts below taken from authentic academic texts. Each contains a phrase with a part cut off. Look at the context and fill in the blanks with the missing half of the words. Sometimes only one letter of a word is missing; sometimes several letters of a word are missing; and sometimes a word will be provided. Use your best handwriting to fill in the blanks. You have 30 minutes to complete this task.

#### Example:

- Mind is **a**                      **o**                      operations carried out by the brain.
  - Mind is **a** set                      **of**                      operations carried out by the brain.
1. Meanwhile, the Arctic Circle has become a hotly contested region; **acc**                       
**t**                      **th**                      U.S. Geological Survey, it holds 13 percent of the world's remaining oil reserves and 30 percent of undiscovered but technically recoverable natural gas deposits.
  2. Both the United States and China **ha**                      **th**                      **sa**                      number of embassies in Africa--forty-eight each.
  3. Because of previous experiences and present perceptions, individuals may perceive themselves in ways **diff**                      **fr**                      **t**                      ways others see them.
  4. After completing each task in both **pa**                      **o**                      **th**                      study, the participants were asked to rate their confidence, satisfaction, and frustration on a 7-point Likert scale.
  5. Due to a growing population and increased consumption, **th**                      **am**                       
**o**                      solid waste generated in Malaysia increased from 16,200 tons per day in 2001 to about 17,000 tons in 2007.
  6. Resources are one of two essential components for change at a small state college. In fact, without new resources in the form of money, time and energy, **t**                      **pro**                       
**o**                      change would have been impossible to carry out.

7. Teachers who work in state schools come face to face with more misbehaviors than the teachers who work in private schools. This may be **th** \_\_\_\_\_ **res** \_\_\_\_\_ **o** \_\_\_\_\_ the home environment that students come from.
8. Recent studies reveal that immigration is going to continue to challenge our schools. Many of the schools are not adequately prepared to assist students with limited English proficiency. This paper discusses **th** \_\_\_\_\_ **pro** \_\_\_\_\_ **o** \_\_\_\_\_ immigrant students who are English limited, and how currently available translator programs can be used to help them.
9. People like Hanny van Arkel are often **ref** \_\_\_\_\_ **t** \_\_\_\_\_ **a** \_\_\_\_\_ amateur astronomers.
10. Talking to peers is easier **be** \_\_\_\_\_ **i** \_\_\_\_\_ **i** \_\_\_\_\_ based on equality, a condition that is less common in conversations with adults.
11. For ease of interpretation, most findings were presented **i** \_\_\_\_\_ **t** \_\_\_\_\_ **fo** \_\_\_\_\_ **o** \_\_\_\_\_ percentage and descriptive statistics.
12. Coaching elite athletes is not **th** \_\_\_\_\_ **sa** \_\_\_\_\_ **a** \_\_\_\_\_ coaching recreational youth athletes.
13. The rationale, according to Friedman, is that when a country reaches **th** \_\_\_\_\_ **le** \_\_\_\_\_ **o** \_\_\_\_\_ economic development required to support a McDonald's, people in that country will stop fighting wars for fear of the resultant economic and personal losses.
14. Such analyses were always used to compare two or more treatments **i** \_\_\_\_\_ **or** \_\_\_\_\_ **t** \_\_\_\_\_ demonstrate one's superiority.
15. Particularly in rural areas in Zimbabwe, women are **li** \_\_\_\_\_ **t** \_\_\_\_\_ **b** \_\_\_\_\_ dependent on their guardians -- either male relatives or, after marriage, their husbands-who speak on their behalf.
16. Within a subsystem **th** \_\_\_\_\_ **a** \_\_\_\_\_ **thr** \_\_\_\_\_ systems - social, cultural, and personality - all of which are interrelated to one another and to the larger systems of society, nations, or to the global village - the world.
17. If more students begin behaving like consumers, colleges and universities may start providing information that affords greater insight into **th** \_\_\_\_\_ **va** \_\_\_\_\_ **o** \_\_\_\_\_ **th** \_\_\_\_\_ education they offer.
18. These instruments measure intelligence through **a** **se** \_\_\_\_\_ **o** \_\_\_\_\_ subtests grouped into a "verbal" and a "performance" scale.
19. Approximately half of those interviewed pointed **o** \_\_\_\_\_ **th** \_\_\_\_\_ **th** \_\_\_\_\_ Ethiopian flag colors are now universally considered the colors of Africa and representative of an African identity.
20. Adults fail to appreciate that teasing can be quite stressful for the child and do not understand that embarrassment in **th** \_\_\_\_\_ **pre** \_\_\_\_\_ **o** \_\_\_\_\_ peers is a major fear of elementary-age children.
21. If you were in a darkroom, **th** \_\_\_\_\_ **wo** \_\_\_\_\_ **b** \_\_\_\_\_ the beginning of the period of trial and error. Expose the paper, process, evaluate. Too light. Repeat. Too dark. Quite time consuming! In Photoshop, our feedback is immediate.
22. Although Alexis de Tocqueville visited America only thirty years before the Civil War, he incorrectly predicted a murderous race war between whites and blacks. He never anticipated that **th** \_\_\_\_\_ **qu** \_\_\_\_\_ **o** \_\_\_\_\_ slavery would be resolved on battlefields where whites would kill whites.
23. In 2003, Congress passed the Syrian Accountability and Lebanese Sovereignty Restoration Act . . . In addition, **i** \_\_\_\_\_ **res** \_\_\_\_\_ **t** \_\_\_\_\_ the Syria accountability act of 2003, Bush issued an order implementing this legislation, which restricted further trade between the two countries and prohibited Syrian aircraft from landing in the United States.
24. Where in our brain do we keep our ABCs? How does our brain provide us with **th** \_\_\_\_\_ **us** \_\_\_\_\_ **o** \_\_\_\_\_ alphabetic characters without thought?
25. In 2006, 1 064 160 people, of whom 130 997 (12.3%) were children and teenagers below the age of 18, were tried in courts for a variety of reasons (ASIGM, 2006a). As seen from these numbers, **th** \_\_\_\_\_ **ra** \_\_\_\_\_ **o** \_\_\_\_\_ juvenile delinquency is notably high.
26. **Th** \_\_\_\_\_ **a** \_\_\_\_\_ **sev** \_\_\_\_\_ reports currently available that predict the use of the cell phone in learning.

27. Divorce and separation is a common feature in the lives of **a la** \_\_\_\_\_ **nu** \_\_\_\_\_ **o** \_\_\_\_\_ children. Most are able to cope relatively well with their existing support networks. However, because it is such a common feature of family life, care should be taken not to minimize the pain and distress it causes many children and young people.
28. If students are enrolling in several traditional courses on campus, they **a** \_\_\_\_\_ **li** \_\_\_\_\_ **t** \_\_\_\_\_ adapt to different learning styles in order to accommodate their respective instructors.
29. The goal of environmental education, may be achieved with a team of well trained, dedicated, religious, socially and environmentally literate teachers. Environmental literacy is an important issue not only for environmental education but for education **a** \_\_\_\_\_ **a wh** \_\_\_\_\_.
30. The importance of interacting with peer's peaks at middle adolescence. **A** \_\_\_\_\_ **th** \_\_\_\_\_ **st** \_\_\_\_\_, adolescents value group over family identity (Busen, 2001).
31. The data are preliminary and with the small sample size, **i** \_\_\_\_\_ **i** \_\_\_\_\_ **n** \_\_\_\_\_ **po** \_\_\_\_\_ **t** \_\_\_\_\_ make definitive statements about the relative performance of the strategies.
32. Second, the parental focus of the research is of value to educators because parent involvement has **be** \_\_\_\_\_ **sh** \_\_\_\_\_ **t** \_\_\_\_\_ influence students' positive self-confidence, self-esteem, and academic success
33. According to instructors, web-based courses offer students more flexibility and control over when and where to participate (Ostiguy and Haffer, 2001), which can lead to greater motivation for students to excel (St. Clair, 1999). Learning in web-based courses **ca** \_\_\_\_\_ **als** \_\_\_\_\_ **b** \_\_\_\_\_ more active (Hacker and Niederhauser, 2000), more student-centered (Sanders, 2001) than taking notes in traditional, passive lectures, and can encourage students to learn in different ways (Yazon et al., 2002).
34. This listening unit is organized into five events. Each of these events is a stage of the unit and may require between one and four class sessions, **de** \_\_\_\_\_ **o** \_\_\_\_\_ **th** \_\_\_\_\_ amount of time a teacher chooses to focus on jazz and jazz artists
35. Researchers comparing men's and women's dissatisfaction with their weight need **t** \_\_\_\_\_ **ta** \_\_\_\_\_ **in** \_\_\_\_\_ **acc** \_\_\_\_\_ the direction of the dissatisfaction. Dissatisfaction with body image in women is normally shown by their desire to lose weight, whereas as many men want to gain weight as lose it.
36. We [professors] must focus on the long term and satisfy ourselves at the moment with the knowledge that one day, hopefully, our students will come to appreciate our efforts on their behalf. It is **fo** \_\_\_\_\_ **th** \_\_\_\_\_ **rea** \_\_\_\_\_ that we should not be overwhelmingly concerned with how our present students feel about us.
37. A student loves nothing better than to catch **hi** \_\_\_\_\_ **or h** \_\_\_\_\_ teacher messing up.
38. It is important to remember that parents generally desire more, as opposed to less, information about their child, **ev** \_\_\_\_\_ **i** \_\_\_\_\_ **the** \_\_\_\_\_ **a** \_\_\_\_\_ unable to articulate relevant questions (Pain, 1999; Quine & Pahl, 1986; Quine & Rutter, 1994).
39. The new Recycle Plus program enabled residents to subscribe to the smallest garbage cart and experience the lowest rates by recycling more. **A** \_\_\_\_\_ **a re** \_\_\_\_\_ **o** \_\_\_\_\_ **the** new system and increased education, yard trimmings recycling increased by 45 percent in one year,
40. Students can be involved in reflecting on their own progress and development **i** \_\_\_\_\_ **a nu** \_\_\_\_\_ **o** \_\_\_\_\_ ways - from journaling to rubric design to recorded assessment.
41. Once students become familiar and comfortable with the process they will, **i** \_\_\_\_\_ **mo** \_\_\_\_\_ **ca** \_\_\_\_\_, react passionately to various musical excerpts, eagerly convey their personal reactions in writing, and have much to say about what they are listening to.
42. The student's level of logic, vocabulary, interest, and maturity will help in determining how to organize your explanation. Broad topics must be simplified **s** \_\_\_\_\_ **th** \_\_\_\_\_ **th** \_\_\_\_\_ scope and detail of the topic are not overwhelming . . .
43. In some experiments in developing countries **i** \_\_\_\_\_ **i** \_\_\_\_\_ **diff** \_\_\_\_\_ for patients to refuse to participate -- one's only chance of receiving any treatment may be as " participant " in a trial (Nuffield Council on Bioethics, 2002, 2005)



44. Research findings indicate that specific instructional strategies, **su** \_\_\_\_\_ **a** \_\_\_\_\_ **th** \_\_\_\_\_ use of real-world examples and independent learning activities, are positively related to mathematics achievement.
45. Education in the American society **nee** \_\_\_\_\_ **t** \_\_\_\_\_ **b** \_\_\_\_\_ supported by the legal institution.
46. What changes are feasible? What changes are essential? **No** \_\_\_\_\_ **o** \_\_\_\_\_ **th** \_\_\_\_\_ questions are easy to answer -- especially the last two.
47. For example, while in one state geography is a required course in order to get a high school diploma, **o** \_\_\_\_\_ **th** \_\_\_\_\_ **o** \_\_\_\_\_ **ha** \_\_\_\_\_, there is not such an obligation in another state.
48. Approximately half of the participants also viewed family as **t** \_\_\_\_\_ **mo** \_\_\_\_\_ **im** \_\_\_\_\_ foundation for their children's life and tried to establish activities that all family members can participate in regularly.
49. In this paper, the author examines **wh** \_\_\_\_\_ **o** \_\_\_\_\_ **n** \_\_\_\_\_ students in early childhood education from a community college receive adequate gender-sensitivity training.
50. A review of the literature showed that **the** \_\_\_\_\_ **ar** \_\_\_\_\_ **n** \_\_\_\_\_ inclusive and detailed criteria for universal software design, although partial lists are available from some sources.
51. The orientation session was also used as an opportunity **t** \_\_\_\_\_ **en** \_\_\_\_\_ **tha** \_\_\_\_\_ all participants understood the intended meaning of the directions.

## APPENDIX F

### THE C-TEST ON PRODUCTION OF THE TOPIC-INDUCED FORMULAIC SEQUENCES

Student's name \_\_\_\_\_

#### Fill- in exercise

**Directions:** Read the sentences below. Each contains a phrase with a part cut off. The phrases are on the topic of *Gray Wolves*. Look at the context and fill in the blanks with the missing part of the words. Sometimes only one letter of a word is missing; sometimes several letters of a word are missing; and sometimes a word will be provided. Use your best handwriting to fill in the blanks. You have up to 10 minutes to complete this task.

1. People today generally **as** \_\_\_\_\_ **wolves** **wi** \_\_\_\_\_ wilderness areas.
2. John James Audubon, a noted naturalist documented several **wolf** **at** \_\_\_\_\_ **o** \_\_\_\_\_ **humans** one of which involves two men traveling through a part of Kentucky near the Ohio border. Both men were severely wounded.
3. Many ranchers insure their sheep and cattle. The organization called "Defenders of Wildlife" does not **com** \_\_\_\_\_ **f** \_\_\_\_\_ **livestock** protected through insurance or a governmental reimbursement programs.
4. Some public officials may support legislation that would **al** \_\_\_\_\_ **farmers** **t** \_\_\_\_\_ **pr** \_\_\_\_\_ their sheep and cows from the predators.
5. The power to list **thr** \_\_\_\_\_ **sp** \_\_\_\_\_ belongs to the Secretary of the Interior and the Secretary of Commerce, who have delegated that power to the Fish and Wildlife Service (FWS).
6. More full-scale **rein** \_\_\_\_\_ **ef** \_\_\_\_\_ **o** \_\_\_\_\_ **gray** **wolf** need to be implemented to prevent the extinction of this species.
7. Conservation officials said they work with the ranchers to try to **ke** \_\_\_\_\_ **th** \_\_\_\_\_ **wolves** **aw** \_\_\_\_\_ **f** \_\_\_\_\_ **humans** and they attach radio collars to the animals to track their whereabouts.
8. Some people would like to know where these carnivores (i.e., meat-eating animals) find food and whether they **po** \_\_\_\_\_ **a** **da** \_\_\_\_\_ **t** \_\_\_\_\_ their pets.
9. The **sla** \_\_\_\_\_ **o** \_\_\_\_\_ **gray** **wolves** began last month, at which time representatives from many animal welfare organizations held protests and met with senators to express their concerns.
10. The U.S. Fish and Wildlife Service estimated the **gray** **wolf** **rei** \_\_\_\_\_ **pr** \_\_\_\_\_, whose goal is to restore the animal to the wild, would cost over \$7,000,000 over nine years.

## APPENDIX G

### IN-CLASS ARGUMENTATIVE ESSAY

**Directions:** You have 40 minutes to respond to the following writing prompt:

Some people agree with Thomas Atwood, the President of America’s National Council for Adoption, who states: “National boundaries should not prevent abandoned children from having families.” Others take the position that orphaned children should remain in their home countries.

**What is your stand on the issue of international adoption? Should a country allow international adoptions or limit adoptions to domestic adoptions only?** Write an essay in which you try to incorporate the following:

- An introductory paragraph that presents both sides of the argument and clearly states your thesis.
- Supporting paragraphs that develop your argument. To support your point of view, refer to the readings you completed prior to class.
- A conclusion that reinforces the position you have taken.

You may write your essay on this paper or your own paper.

## APPENDIX H

### ESSAY RATING RUBRIC

Essay Rating Rubric			
	Score	Level	Criteria
Content	x 7.5	4	<b>EXCELLENT TO VERY GOOD: knowledgeable</b> – substantive – through development of thesis - relevant to assigned topic
		3	<b>GOOD TO AVERAGE: some knowledge of the subject</b> – adequate substance – some development of thesis evident – mostly relevant to topic, but lacks detail
		2	<b>FAIR TO POOR: limited knowledge of subject</b> – little substance – inadequate development of thesis – limited relevance to topic
		1	<b>VERY POOR: does not show knowledge of subject</b> - non substantive – no development of thesis OR not enough prose to evaluate
Organization	x 5	4	<b>EXCELLENT TO VERY GOOD: well-organized</b> - ideas clearly stated and supported – paragraph structure is coherent and unified
		3	<b>GOOD TO AVERAGE: Organization may be uneven</b> – ideas loosely organized but main ideas stand out – limited support – only minor lapses in paragraph structure and coherence
		2	<b>FAIR TO POOR: ideas confused or disconnected</b> – ideas lack logical sequencing and development – issues with paragraph structure and coherence
		1	<b>VERY POOR: lacks organization</b> – prose does not communicate ideas OR there is not enough prose to evaluate
Vocabulary	x 5	4	<b>EXCELLENT TO VERY GOOD: sophisticated range</b> – effective word or idiom choice and usage – very good word form mastery – appropriate register with only minor and infrequent issues - meaning clear
		3	<b>GOOD TO AVERAGE: adequate range</b> – occasional errors in word or idiom choice and usage – some word form mastery - meaning generally clear – appropriate register with some issues
		2	<b>FAIR TO POOR: limited range</b> – frequent errors in word or idiom choice and usage – little word form mastery – meaning frequently confused or obscured – limited inappropriate register and frequent errors
		1	<b>VERY POOR: no range</b> – errors in word or idiom choice and usage dominate – meaning confused OR not enough prose to evaluate
Language Use	x 6.25	4	<b>EXCELLENT TO VERY GOOD: Effective complex constructions</b> – few errors of agreement, tense, number, word order, articles, pronouns, prepositions
		3	<b>GOOD TO AVERAGE: Minor errors in complex constructions</b> – several errors of agreement, tense, number, word order, articles, pronouns, prepositions
		2	<b>FAIR TO POOR: major errors in complex constructions</b> – frequent errors of agreement, tense, number, word order, articles, pronouns, prepositions
		1	<b>VERY POOR: no mastery of sentence construction rules</b> – dominated by errors OR not enough prose to evaluate
Mechanics	x 1.5	4	<b>EXCELLENT TO VERY GOOD: demonstrates mastery of conventions</b> – few errors of spelling, punctuation, capitalization, paragraphing
		3	<b>GOOD TO AVERAGE: adequate mastery of conventions</b> - occasional errors of spelling, punctuation, capitalization, paragraphing
		2	<b>FAIR TO POOR: little mastery of conventions</b> - frequent errors of spelling, punctuation, capitalization, paragraphing
		1	<b>VERY POOR: no mastery of conventions</b> - dominated by errors of spelling, punctuation, capitalization, paragraphing OR not enough prose to evaluate

## APPENDIX I

### INDEPENDENT SAMPLES TEST RESULTS

Measure	Levene's Test of Equality of Variances		t-test for Equality of Means		
	<i>F</i>	Sig.	<i>t</i>	<i>DF</i>	Sig.
C-test AFS Pretest					
Equal variances assumed	.087	.769	<b>-.192</b>	<b>58</b>	<b>.848</b>
Equal variances not assumed			-.200	55.171	.842
C-test AFS Posttest					
Equal variances assumed	8.173	.006	-4.777	55	.000
Equal variances not assumed			<b>-5.276</b>	<b>53.793</b>	<b>.000</b>
AFS Essay 1					
Equal variances assumed	1.923	.171	<b>-1.426</b>	<b>53</b>	<b>.160</b>
Equal variances not assumed			-1.543	51.811	.129
AFS Essay2					
Equal variances assumed	.207	.651	<b>1.323</b>	<b>49</b>	<b>.192</b>
Equal variances not assumed			1.366	44.804	.179
C-test T-I Posttest					
Equal variances assumed	3.708	.059	-7.642	56	.000
Equal variances not assumed			<b>-7.927</b>	<b>52.591</b>	<b>.000</b>
Essay T-I Pretest					
Equal variances assumed	.107	.745	<b>.184</b>	<b>61</b>	<b>.855</b>
Equal variances not assumed			.184	49.338	.855
Essay T-I Pretest					
Equal variances assumed	10.473	.002	-2.977	52	.004
Equal variances not assumed			<b>-3.686</b>	<b>49.759</b>	<b>.001</b>
Writing Pretest					
Equal variances assumed	2.520	.118	<b>.212</b>	<b>60</b>	<b>.833</b>
Equal variances not assumed			.223	56.627	.825
Writing Pretest					
Equal variances assumed	.001	.981	<b>-.110</b>	<b>49</b>	<b>.913</b>
Equal variances not assumed			-.110	40.232	.913

Note. C-test AFS = production of academic formulaic sequences in a controlled situation (i.e., C-test); Essay AFS = production of academic formulaic sequences in an uncontrolled situation (i.e., essay); C-test T-IFS production of topic-induced formulaic sequences in a controlled situation (i.e., C-test); Essay T-IFS = production of topic-induced formulaic sequences in an uncontrolled situation (i.e., essay); Writing = in-class essay

## REFERENCES

- Altenberg, B. (2002). Causative constructions in English and Swedish: A corpus-based contrastive study in B. Alenberg and S. Granger (Eds.), *Lexis in contrast: Corpus-based approaches* (pp. 97-118). Amsterdam, PH: John Benjamins Publishing Company.
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22, 173-195.
- Bachman, L., & Palmer, A. (2010). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Barkaoui, K. (2010). Do ESL Essay Raters' Evaluation Criteria change with experience? A mixed-methods, Cross-Sectional study, *TESOL Quarterly*, 44, 1, 31-57.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263-86.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard and S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181-190). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371-405.
- Biber, D., Johansson, D., Leech, G., Conrad, S., & E. Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum* 1(1), 1-14.
- Christison, M. A. (forthcoming). *The practice of content-based Instruction*. Ann Arbor, MI: University of Michigan Press.
- Coady, J. (1997). L2 vocabulary acquisition: A synthesis of the research. In J. Coady & T. Huckling (Eds.), *Second language vocabulary acquisition* (pp. 273-290). Cambridge: Cambridge University Press.
- Elly, W., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly*, 19, 53-67.

- Cobb, T. (2010). N-gram Phrase Extractor (Version 4) Retrieved from <http://lextutor.ca/tuples/eng/>
- Cobb, T. (2010). Text Lex Compare (Version 2.2). Retrieved from [http://www.lextutor.ca/text\\_lex\\_compare/](http://www.lextutor.ca/text_lex_compare/)
- Corson, D. (1985). *The lexical bar*. Oxford: Pergamon Press.
- Corson, D. J. (1995). *Using English words*. Dordrecht: Kluwer Academic Publishers.
- Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. Fitzmaurice, and D. Biber (Eds.), *Using corpora to explore linguistic variation*. Amsterdam, Ph: John Benjamins Publishing Company.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397–423.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17, 391-406.
- Cowie, A.P. (1992). Multiword lexical units and communicative language teaching. In Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics*. London: MacMillan.
- Coxhead, A. (2000). A new academic word list, *TESOL Quarterly*, 34, 213-238.
- Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In F. Meunier and S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 149-162). Amsterdam, PA: John Benjamins Publishing Company.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, 129–147.
- Coxhead, A., Bunting, J., Byrd, P., & Moran, K. (forthcoming). *The academic word list: Collocations and recurrent phrases*. Ann Arbor: University of Michigan Press.
- Davies, M. (2008). Corpus of Contemporary American English (COCA) available at <http://www.americancorpus.org/>
- Davis, K. (1995). Qualitative theory and methods in applied linguistics research. *TESOL Quarterly*, 29, 427-453.

- De Cock, S. (2003). Recurrent sequences of words in native speaker and advanced learner spoken and written English: A corpus-driven approach. Unpublished PhD dissertation. Louvain-la-Neuve: Université catholique de Louvain.
- Dörnyei, Z. & Csizér, K. (2002). Some dynamics of language attitudes and motivation: Results of longitudinal nationwide survey. *Applied Linguistics*, 23, 421-462.
- Dörnyei, Z. (2002). The motivational basis of language learning tasks. In P. Robinson (Ed.), *Individual differences and instructed language learning*, (pp. 137-158). Amsterdam: John Benjamins Publishing.
- Dörnyei, Z., Durow, V., Zhan, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt, (Ed.), *Formulaic sequences* (pp. 87-106), Philadelphia: John Benjamins Publishing.
- Dörnyei, Z., Durow, V., & Zhan, K. (2004). Individual differences and their effects on formulaic sequences acquisition. In Schmitt N. (2004). *Formulaic sequences* (pp. 87-106). Amsterdam: John Benjamins Publishing.
- Duin, A. H., & Graves, M. F. (1987). Intensive vocabulary instruction as a prewriting technique, *Reading Research Quarterly*, 22, 311-330.
- Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistic, corpus linguistics, and TESOL. *TESOL Quarterly*, 42, 375-396.
- Ellis, R. (2009). *The study of second language acquisition*. Oxford: Oxford University Press.
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions, *Journal of Second Language Writing*, 4, 139-55.
- Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In R. Corrigan, E. A. Moravcsik, H. Ouali, and K.M. Wheatley (Eds.), *Formulaic Language, Volume 2*, (pp. 323-346) Philadelphia: John Benjamins Publishing Company.
- Ferris, D. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, 28, 414-420.
- Firth, J. (1957). *Papers in linguistics: 1934-1951*. London: Oxford University Press.
- Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40, 273-29.



- Folse, K. (2008). Myth 1: Teaching vocabulary is not the writing teacher's job. In J. Reid (Ed.), *Writing myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press.
- Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 125-144). Oxford: Clarendon Press.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing. Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 146-160). Oxford: University Press.
- Grobe, C. H. (1981). Syntactic maturity, mechanics and vocabulary as predictors of quality rating. *Research in the Teaching of English, 15*, 75-85.
- Harley, B., & King, M. L. (1989). Verb lexis in the written compositions of young L2 learners, *Studies in Second Language Acquisition, 11*, 415-439.
- Hatch, J. (2002). *Doing qualitative research in education settings*. Albany: State University of New York Press.
- Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A.P.Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 161-187). Oxford: Clarendon Press.
- Hulstijn, J. H. (2001). Intentional and incidental vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *In cognition and second language instruction*, (258-286), Cambridge: Cambridge University Press.
- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16*, 148-164.
- Hyland, K. (2008). Lexical clusters: Text patterning in published and post-graduate writing, *International Journal of Applied Linguistics, 18*, 41-61.
- Jacobs, Harfield, Hughey, & Wormeth (1981). *Testing ESL composition: A practical approach*. Boston: Newbury House.
- Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences. In N. Schmitt, (Ed.), *Formulaic sequences* (pp. 269-300), Philadelphia: John Benjamins Publishing.

KeyWords Extractor (Version 1)

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, 73, 4, 440-463.

Laufer, B. & Shmueli, K. (1997). Memorizing new words: does teaching have anything to do with it? *RELC Journal* 28, 1, 89-108.

Leki, I., & Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the discipline. *TESOL Quarterly*, 28, 81-101.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Lang Writing*, 18, 85-102.

Linnamd, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Malmö, Sweden: Liber Fölag Malmö.

Martinez, R. and Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33, 229-320.

McClure, E. (1991). A comparison of lexical strategies in L1 and L2 written English narratives. *Pragmatics and Language Learning*, 2, 141-154.

Meunier, F. & Granger S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins Publishing Company.

Moon, R. (1998). Frequencies and forms of phrasal lexemes in English. In A.P.Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 79-100). Oxford: Clarendon Press.

Murray, D. E., & Christison, M. A. (2010). *What Teachers Need to Know II: Facilitating Learning*. New York: Routledge/Taylor & Francis.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2005). Teaching and learning vocabulary. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 581-596). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Nation, I.S.P. & Gu, P. Y. (2007). *Focus on vocabulary*. Sydney, Australia: National Center for English Language Teaching and Research Macquarie University.

- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*, Oxford: Oxford University Press.
- Neff van Aertselaer, J. (2008). Contrasting English-Spanish interpersonal discourse phrases in phraseology. In F. Meunier and S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 85-100). Amsterdam, PA: John Benjamins Publishing Company.
- Neff, J., Ballesteros, F., Dafouz, E., Martínez, F., Rica, J. P., Diez, M., & Prieto, R. (2004). The expressions of writer stance in native and non-native argumentative texts. In R. Facchinetti and F. Palmer (Eds.) *English modality in perspective. genre analysis and contrastive studies* (pp. 141-161). Frankfurt: Peter Lang.
- Nielsen, L., & Piche, G. L. (1981). The influence of headed nominal complexity and lexical choice on teachers' evaluation of writing. *Research in the Teaching of English*, 15, 65-73.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.
- Ohlrogge, A. (2009). Formulaic expressions in intermediate EFL writing assessment. In R. Corrigan, E. A. Moravcsik, H. Ouali, and K.M. Wheatley (Eds.) *Formulaic language, Volume 2*, (pp. 375-386) Philadelphia: John Benjamins Publishing Company.
- Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier and S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 110-120). Amsterdam, PA: John Benjamins Publishing Company.
- Paribakht, T., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.) *Second language vocabulary acquisition* (pp. 174-200). Cambridge: Cambridge University Press.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Beverly Hills, CA: Sage.
- Patton, M. Q. (1990). *Qualitative evaluation and Research methods* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications.
- R. Corrigan, E. A. Moravcsik, H. Ouali, and K.M. Wheatley (Eds.) *Formulaic language, Volume 2*, (pp. 375-386) Philadelphia: John Benjamins Publishing Company.
- Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly*, 19, 229-258.

- Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students, *TESOL Quarterly*, 22, 69-90.
- Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction*. (pp.3-32) Cambridge: Cambridge University Press.
- Schmitt, N. (2000). *Vocabulary instruction in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins Publishing Company.
- Schmitt, N., Dorneyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt, (Ed.), *Formulaic sequences* (pp. 55-86), Philadelphia: John Benjamins Publishing.
- Scott, M., & Tribble, C. (2006). English for academic purposes: Building an account of expert and apprentice performances in literary criticism. In M. Scott and C. Tribble (Eds.), *Textual patterns: key words and corpus analysis in language education* (pp. 131-159). Amsterdam, PA: John Benjamins Publishing Company.
- Simpson-Vlach, R., & Ellis, C. N. (2010). An academic formulas list. New methods in phraseology research. *Applied Linguistics*, 31, 487-512.
- Sinclair, J. 1991. *Corpus, concordance, collocation: Describing English language*. Oxford: Oxford University Press.
- Song, B. & Caruso, I. (1996). Do English and ESL faculty differ in evaluating the essays of native-English speaking and ESL students? *Journal of Second language Writing*, 5, 163-182.
- Swinney, D & Cutler, A. (1979). The access and processing of idiomatic expressions. *JVLVB*, 18, 523-534.
- Wajnryb, (1988). The dicto-gloss method of language teaching: A communicative approach to teaching grammar. *English Teaching Forum*, 26(3); 35-38.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Woolard, (2000). Collocation--encouraging learner independence. In M. Lewis (ed.) *Teaching collocation: Further development in the lexical approach* (pp. 28-46), Boston: Heinle.

Yorio, C. A. (1989). Idiomatic as an indicator of second language proficiency. In K. Hyltenstam & L.K. Obler (Eds), *Bilingualism across the lifespan* (pp.55-72), Cambridge: Cambridge University Press.

Zimmerman, C.B. (1997) Do reading and interactive vocabulary instruction make a difference? An empirical study. *TESOL Quarterly*, 31, 121-140.

