

2014-01-01

A Corpus-Based Study of the Use of Prepositional Verbs in Second Language Emergent Academic Writing

Elizabeth Wilcoxon

University of Texas at El Paso, wilcoxon.elizabeth@gmail.com

Follow this and additional works at: https://digitalcommons.utep.edu/open_etd



Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [English Language and Literature Commons](#), and the [Linguistics Commons](#)

Recommended Citation

Wilcoxon, Elizabeth, "A Corpus-Based Study of the Use of Prepositional Verbs in Second Language Emergent Academic Writing" (2014). *Open Access Theses & Dissertations*. 1758.
https://digitalcommons.utep.edu/open_etd/1758

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

A CORPUS-BASED STUDY OF THE USE OF PREPOSITIONAL VERBS
IN SECOND LANGUAGE EMERGENT ACADEMIC WRITING

ELIZABETH MARIE WILCOXON

Department of Languages and Linguistics

APPROVED:

Alfredo Urzúa, Ph.D., Chair

Natalia Mazzaro, Ph.D.

Alberto Esquinca, Ph.D.

Bess Sirmon-Taylor, Ph.D.
Interim Dean of the Graduate School

Copyright ©

by

Elizabeth M. Wilcoxon

2014

Dedication

To my loving husband Aaron, who was serving our country overseas during the inception and production of this project. He never ceases to support me in any endeavor I set my mind to, even from 7,382 miles away. I love you.

A CORPUS-BASED STUDY OF THE USE OF PREPOSITIONAL VERBS
IN SECOND LANGUAGE EMERGENT ACADEMIC WRITING

by

ELIZABETH MARIE WILCOXON, BA

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF ARTS

Department of Languages and Linguistics

THE UNIVERSITY OF TEXAS AT EL PASO

May 2014

Acknowledgements

I would like to extend a million thanks to the faculty and staff in the department of Languages and Linguistics at the University of Texas at El Paso. This entire UTEP experience has been incredible for me because of the supportive family-like atmosphere on our few floors of the Liberal Arts building. I would also like to thank the faculty in our department of whom I have been a student or with whom I have worked. Their expertise in our field has had an invaluable impact on me. I would also like to acknowledge Dr. Natalia Mazzaro and Dr. Alberto Esquinca, members of my thesis committee. I am thankful for the time they spent to read my paper and give constructive feedback.

I would like to personally thank Dr. Alfredo Urzúa, the chair of my committee and supervisor in my current position as a graduate teaching assistant. From the first class I took with him, it was very evident to me that his knowledge, compassion, and professionalism have made him one of the best, if not *the* best teacher that I know. The process of writing this thesis from a brainstorm on a scrap piece of paper to the final product has been seamless and quite enjoyable thanks to his guidance and encouragement.

Last but certainly not least, I would like to express my gratitude to my dear friend Sevda Rahymov, who has shared classes, group projects, and many cups of coffee with me. I treasure her encouragement throughout the course of our graduate program, and the friendship we have forged is precious to me.

Table of Contents

| | |
|---|------|
| Acknowledgements..... | v |
| Table of Contents..... | vi |
| List of Tables | viii |
| List of Figures..... | ix |
| Introduction..... | 1 |
| Chapter 1: Literature Review..... | |
| 1.1 A brief history of corpus linguistics | 3 |
| 1.2 Learner corpora and second language studies | 4 |
| 1.3 Second language acquisition and interlanguage | 5 |
| 1.4 The umbrella of multi-word Items..... | 6 |
| 1.5 Prepositional verbs (PVs) | 7 |
| 1.5.1 Definition of PVs..... | 7 |
| 1.5.1.1 How to differentiate PVs from phrasal verbs | 9 |
| 1.5.1.2 How to differentiate PVs from verb plus prepositional phrases..... | 11 |
| 1.6 Prepositional verbs in the literature | 11 |
| 1.6.1 Prepositional verbs and language acquisition..... | 11 |
| 1.6.2 Prepositional verbs in corpora | 12 |
| 1.6.3 Errors in preposition selection | 13 |
| 1.6.3.1 Reasons for errors | 15 |
| 1.7 Previous methods of analysis..... | 15 |
| Chapter 2: Methodology | |
| 2.1 Research questions..... | 17 |
| 2.2 UTEP Learner Corpus of Academic English (ULCAE)..... | 17 |
| 2.2.1 Description..... | 17 |
| 2.2.2 Description of data selected from ULCAE..... | 18 |
| 2.2.2.1 Description of writing assignments in ESOL 1610 and 1309 | 18 |
| 2.2.2.2 Description of writing assignments in ESOL 1311 and ESOL 1312 | 19 |
| 2.3 Method of analysis..... | 20 |
| 2.4 Data -- Sub-Corpus A | 22 |
| 2.4.1 ESOL 1610 and ESOL 1309..... | 22 |

| | |
|---|----|
| 2.4.2 ESOL 1311 and ESOL 1312..... | 23 |
| Chapter 3: Results..... | |
| 3.1 Frequency of PVs | 25 |
| 3.2 Contrasting registers of PV usage..... | 26 |
| 3.2.1 Examples of CONV PVs in ESOL 1610 and ESOL 1309 | 28 |
| 3.2.2 Examples of ACAD PVs in ESOL 1311 and ESOL 1312 | 28 |
| 3.3 Most frequent PVs per course/level..... | 28 |
| 3.3.1 Examples of top two most frequent PVs in ESOL 1610 | 30 |
| 3.3.2 Examples of top two most frequent PVs in ESOL 1309 | 30 |
| 3.3.3 Examples of top two most frequent PVs in ESOL 1311 | 30 |
| 3.3.4 Examples of top two most frequent PVs in ESOL 1312 | 31 |
| 3.4 Correct and Incorrect Usage of PVs | 31 |
| 3.4.1 Examples of incorrect usage of PVs in ESOL 1610 and ESOL 1309 | 32 |
| 3.4.2 Examples of incorrect usage of PVs in ESOL 1311 and ESOL 1312 | 34 |
| Chapter 4: Discussion and Implications | |
| 4.1 Discussion..... | 35 |
| 4.2 Limitations | 38 |
| 4.3 Suggestions for further study | 38 |
| References..... | 40 |
| Vita..... | 42 |

List of Tables

| | |
|---|----|
| Table 1.1: PVs according to register..... | 13 |
| Table 1.2: Different types of errors in verb-noun collocations by frequency. | 14 |
| Table 2.1: Basic information for ESOL 1610 and ESOL 1309..... | 23 |
| Table 2.2: Basic information for ESOL 1311 and ESOL 1312..... | 24 |
| Table 3.1: Subcorpus A with normed frequencies..... | 25 |
| Table 3.2 : Types of PVs according to register | 26 |
| Table 3.3: Most frequent PVs per course/level. | 29 |
| Table 3.4: Number and percentage of tokens of PVs used with appropriate prepositions..... | 31 |
| Table 3.5.1: Correct and incorrect usage of most frequent PVs in ESOL 1610 and ESOL 1309. | 32 |
| Table 3.5.2: Correct and incorrect usage of most frequent PVs in ESOL 1311 and ESOL 1312. | 33 |

List of Figures

| | |
|--|----|
| Figure 2.1: Sample extraction of concordancing lines | 21 |
| Figure 3.1: Contrast of CONV and ACAD PVs..... | 27 |

Introduction

In this paper, I report on a corpus-based study of prepositional verbs in L2 emergent academic writing using data collected from English as a Second Language (ESL) and English for Academic Purposes (EAP) classrooms at the University of Texas at El Paso. Prepositional verbs are relatively common in academic prose (lacking the informal overtones of phrasal verbs) and draw on the full set of prepositions (Biber, Johansson, Leech, Conrad, & Finnegan, 1999). They are also difficult to acquire given the challenges posed by prepositional usage for most ESL/EAP learners. The literature in this area highlights the importance of prepositional verbs in L2 learning (Tetreault & Chodorow, 2008; Hong, Rahim, Hua, and Salehuddin, 2011); however, there is a lack of empirical studies of these verbs in ESL and EAP writing.

The motivation for the study comes from my experiences as an ESOL writing tutor and novice post-secondary instructor of ESL/EAP students, mostly Spanish-speaking learners. I noticed that students use prepositional verbs early in their process of English language acquisition with varying degrees of accuracy, depending on the verb used. Using data from the UTEP Learner Corpus of Academic English (ULCAE), a local learner corpus, I explore the use of prepositional verbs by means of frequency analyses aimed at determining the degree to which students use prepositional verbs in their essays, as well as comparing their occurrence across proficiency levels. I discuss these patterns of occurrence and the distribution of conversational and academic prepositional verbs (based on Biber et al., 1999), as well as findings from an error analysis of a sub-sample of the data conducted in order to determine levels of accuracy in the use of prepositional verbs.

The main goal of this study is to provide empirical evidence regarding the use of prepositional verbs in emergent academic writing within the context of an ESL/EAP program. The research questions explored in this project aim to provide specific information regarding both the frequency and accuracy of PVs. In order to investigate the use of these linguistics forms, corpus-based techniques are used.

Sylviane Granger, in the volume *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching* (2002), describes the nature of techniques within corpus linguistics as “particularly powerful methodology,” and these techniques, instead of making other methodologies for analyzing language obsolete, seek to complement them (p. 4).

Chapter 1: Literature Review

1.1 A brief history of corpus linguistics

Corpora, bodies of written or spoken texts, have been studied in the field of linguistics since the late 1980's, and their use in research has only grown in use as technology has advanced. Now more than ever, it is easier to transcribe, compile, and analyze data with technology and software programs designed specifically for corpora usage. Researchers have defined a corpus as a “principled collection of texts” (O’Keeffe, McCarthy, and Carter, 2007), and “authentic language which has been compiled for a particular purpose” (Flowerdew, 2012). While corpus linguistics does not have a long history in comparison with other sub-fields of linguistics, the field itself has experienced many changes over its relatively short life.

Bonelli (2012), in the *Routledge Handbook of Corpus Linguistics*, cites a previous work (Bonelli & Sinclair, 2006) which classifies the history of corpus linguistics into three “generations”: from 1960 to 1980, there was no electronic material and corpora could contain up to a million words; from 1980 to 2000, the development of the scanner allowed researchers to build corpora of up to twenty million words; and 2000 to the present, in which advances in technology have allowed for virtually unlimited amounts of corpus data to be available. Furthermore, the impetus behind corpus linguistics has changed with the times as well, largely due to the rise of technology and the quantities of information available. Corpus linguistics began as a methodology for describing natural language data. The methodology is one of a “bottom-up” approach, so to speak, in which the data (e.g. corpora) are already given and the researcher describes the patterns of usage that already exist. This differs from other empirical approaches in linguistics, in which researchers set out to elicit certain data from their participants.

In the *Routledge Handbook*, Bonelli (2012) describes the move from a methodological approach (mainly qualitative) in corpus linguistics to that of a theoretical one, thanks to a quantitative explosion that has “shaken the underlying assumptions behind many well-established theoretical positions in the

field” (p. 17). In short, the widespread improvements in technology have allowed corpus linguistics to literally explode as a field, and in turn, researchers are now able to observe patterns of language use within vast amounts of data that have not been observed before.

1.2 Corpus analysis and second language studies

While corpus research in general has been growing since the 1960’s, Granger (2002) explains that learner corpus research has only been active since the 1980’s. Before the field started to grow in popularity, corpus linguistics and second language research were not strongly connected at all. However, research in the field of corpus linguistics has proven that techniques used within corpus linguistics can be powerful tools for analysis. One way corpus linguistics can be used is to provide evidence of how speakers use language.

Unlike methodological approaches that elicit data from generally a small number of participants, corpus linguistics uses language as it occurs naturally in speech and writing from many participants at once. Learner corpora are available for qualitative and quantitative analysis, both of which are of prime importance in the world of scientific research. To explain, a learner corpus is built from data collected from learners of a language. In these corpora, researchers are able to observe the language that learners actually produce in natural environments, and thus they can attempt to draw conclusions from a large set of data about such topics as patterns of usage and frequency of occurrence of a particular linguistic item. Most learner corpora are cross-sectional, providing data at one point in time. Longitudinal and quasi-longitudinal data are much less common, as it is difficult and time-consuming to collect these across time. Despite the short history of corpus linguistics, the type of data collected from learner corpora can endeavor to inform classroom teaching and help to improve instructional methods, especially if the data are collected locally.

One of the main uses of learner corpora is in the field of second language pedagogy, as cited by O’Keeffe et al. (2007). With the help of corpus data, applied linguists with the specific interest in second

language education have started to describe how learners acquire and use language. Analyses of learner corpora can have important effects in how second language development is viewed and understood, including classroom language and emergent academic discourse. O’Keeffe et al. (2007) posit that the study of corpora could change “long-held” notions of education and pedagogy. In the same vein, corpora can also bridge the gap between the cognitive science of linguistics and many other areas including sociolinguistics, teaching, grammar, and translation.

Findings from corpora have already begun to bridge this gap. In *How to Use Corpora in Language Teaching* (Sylviane Granger, Joseph Hung, & Stephanie Petch-Tyson, 2004), Amy Tsui outlines some of the prominent studies done in corpus linguistics which have explored the following four areas: lexical collocation, syntactic patterning, genre analysis, and discourse structure. These studies include analyses conducted with corpora such as the Brown Corpus (Kjellmer, 2004) and Carter & McCarthy’s Spoken English Corpus (Carter & McCarthy, 1997).

1.3 Second language acquisition and interlanguage

Before delving into the topic of multiword items and its subset of prepositional verbs, the focus of this study, it is worth mentioning that any learner corpus consists of a representation of learner language or “interlanguage” (Selinker, 1972). Interlanguage (IL) is a system of language used by L2 learners that enables them to approximate the L2, and which reflects the learners’ current hypotheses about the target language. An interlanguage, thus, represents a learners’ systematic and rule-governed use of language, even if non-native-like, with errors reflecting natural processes of language acquisition, such as transfer from the first language or overgeneralization of certain L2 rules (Selinker, 1972). From this perspective, learner language is not viewed as deficient or limited, but as a dynamic, evolving, non-linear language system. The more linguists can understand and make sense of certain aspects of interlanguage, the more informed L2 instruction can be, thus producing more competent and native-like language from learners. IL encompasses a broad range of aspects of learner language, but in this paper I

discuss and explore a very specific lexico-grammatical aspect of IL development in the area of multiword items, and even more specifically, prepositional verbs. The next section focuses on these linguistic forms.

1.4 The umbrella of multiword items

Prepositional verbs (henceforth PVs) fall under a category of lexical items called multiword items (MWIs), sometimes also referred to as multiword expressions (Baldwin, 2005). These items come in many classes to include idioms, phrasal verbs, prepositional verbs, stock phrases, prefabrications (or prefabs), and formulaic sequences, among others. They are formed with two or more words, sometimes up to seven, which are part of the speaker's lexicon that enables him or her to sound more native-like (Gardner & Davies, 2007). Wray (2000) defines formulaic language as:

“A sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is stored and retrieved whole from the memory at the time of use, rather than being subject to generation or analysis by the language grammar” (in Ellis, 2012, p. 26).

Cobb (2003) states that MWIs allow the speaker a two-for-one deal that results in “cognitive economy” and brain power for other things such as discourse planning. In addition, a very important characteristic of MWIs in general is that they are idiosyncratic in their lexicon, syntax, semantics, and pragmatics. For example, idioms are idiosyncratic in that they have a meaning particular to the whole of the phrase, and words cannot be substituted to mean the same thing, such as in “kick the bucket.” One cannot say something such as “kick the pail” and have it mean the same thing pragmatically as “kick the bucket.” These forms are predictable, lexically and syntactically, meaning that the words are always located together (or are collocated) and the meaning is always the same. Taking the previous example of “kick the bucket,” it will always carry the semantic meaning of “to die.”

The acquisition of MWIs enables the learner to speak, read and write more fluently (Hong et al., 2011). On the other hand, the lack of these items in a learner's interlanguage may result in "foreign sounding" speech or writing (Cobb, 2003). Research makes it clear that these items have to be learned, not simply observed and absorbed while learning the language. Gardner and Davies (2007) point out that language teachers know how important it is for their students to grasp the concept of MWIs, but because of the idiosyncratic nature of these items, it is difficult to know how and what to include in second language curricula. Laufer and Waldman (2011) go so far as to say that difficulty with collocations and MWIs does not discriminate between learners with more years of instruction in the L2, particular native language backgrounds, or what they are asked to do in the L2.

1.5 Prepositional verbs (PVs)

As already mentioned, the present study focuses on one specific type of MWIs, i.e., prepositional verbs. One important aspect that needs to be discussed with regards to PVs is how they are similar and different from other MWIs. According to Biber et al. (1999), there are two main lexical verbs which contain multiple words: phrasal verbs (verb plus adverbial particle) and prepositional verbs (verb plus preposition). PVs employ the full set of prepositions (e.g., be used *in*, look *for*, be based *on*, be associated *with*). When these verbs do not contain all their necessary constituents, whether omitted or substituted, it can make the interlanguage of an L2 English learner sound less "native" because he or she is not using the accepted combinations of verbs with their prepositional collocates. The following section provides a general description of what prepositional verbs are as well as examples and tests illustrating their unique properties in order to get a better sense of the differences between PVs and other MWIs.

1.5.1 Definition of PVs

Generally the research agrees on the structure of PVs. De Haan (1988) gives some initial insight into what PVs are, but not a complete syntactical description. First, in PVs, the verb and its following

preposition are “closely related” (p. 121). Biber et al. (1999) provide a clearer definition of a PV, saying that “the verb + preposition functions as a single semantic unit, with a meaning that cannot be derived completely from the individual meanings of the two parts” (p. 414). In addition, the two parts consist of a “PP [prepositional phrase] argument made up of a specified preposition head and NP [noun phrase] argument” (Baldwin, 2005, p. 116). According to this definition, PVs are transitive, taking an object, such as in *refer to (the book)* or *come across (the letter)* (p. 116).

Biber et al. (1999) gives two structural patterns of PVs, both of which are transitive:

Pattern 1 - Noun phrase (NP) + **verb** + preposition + noun phrase (NP)

Her scholarship to college **depended** on her final GPA.

Pattern 2 - NP + **verb** + NP + preposition + NP

That organization **exists to save** humpback whales from extinction.

According to De Haan (1988), one test to identify PVs is by determining if the object of the preposition answers the question *who* or *what*. Refer to the following examples:

- 1) a. He looked at the dog
b. What did he look at?
c. *Where did he look?
- 2) a. She was walking on the sidewalk
b. Where was she walking?
c. *What was she walking on?

In the first example, the object of the preposition answers the question *what*, which makes “look at” a prepositional verb. However, in the second example, the object of the preposition answers the question *where*, which does not meet the criteria for a prepositional verb, and the answer to the question, “at the dog,” is actually a prepositional phrase. Other sources (Biber et al., 1999; Quirk, 1972) include

some of the same tests to determine if a verb plus preposition can indeed be categorized as a prepositional verb.

Baldwin (2005) cites Huddleston and Pullum (2002) in that one characteristic of PVs is that of passivization, in which some PVs cannot be passivized (fixed PVs) and others can (mobile PVs). Using mobile and fixed PVs, Baldwin (2005) provides a passivization test for each that helps distinguish PVs from other similar structures. Consider the following examples of fixed PVs:

3) a. He *came across* a problem

b. *A problem was *come across* by him

4) a. She *walked down* the path

b. *The path was *walked down*

The above fixed PVs cannot be passivized. Baldwin states that the object of the PV must follow directly after the PV. In (3a) and (4a), the verb *come across* and *walked down* respectively can only be followed by their objects. This test concludes that fixed PVs cannot be passivized.

There are also mobile PVs, which can be passivized, and these are seen frequently in academic prose.

5) a. She *referred to* the book in her paper

b. The book was *referred to* in her paper

6) Smoking *was associated with* lung cancer

In (5a) *referred to* can either appear before or after its object. (6a) contains a PV that is acceptable in the passive voice.

The two previous tests provide clarification on the structure and movement of PVs, but there is a more reliable test to follow.

1.5.1.1 How to differentiate PVs from phrasal verbs

A strong and reliable test to determine PVs which also helps differentiate them from phrasal verbs is the ability to insert the object of the preposition between the verb and preposition, which is

described in Quirk (1972). Consider the following phrasal verbs, which allow the grammatical object to be inserted between the verb and the preposition:

7) a. John *called up* the man

b. John called the man up

8) a. She left out that part in her paper

b. She left that part in her paper out

In contrast to examples (7) and (8), in the following PVs, the object cannot be inserted between the verb and the preposition, as it is part of the prepositional phrase (PP).

9) a. John *called on* the man

b. *John called the man on

10) a. She *put up* with the heat

b. *She put with the heat up

Another strong test used to differentiate PVs from phrasal verbs is that of adverb intrusion, cited by Teschner and Evans (2000).

11) a. They called *frequently* on their teacher.

b. They called on their teacher *frequently*.

12) a. They *frequently* called up their teacher.

b. *They called *frequently* up their teacher.

In (11), *called on* is a PV and accepts the adverb between its constituents. However, in (12) the adverb cannot be inserted between *called up* because it is a phrasal verb, not a PV.

Finally, Teschner and Evans (2000) employ the wh-word fronting test to distinguish between PVs and phrasal verbs.

13) a. Which woman did they call on?

b. On which woman did they call?

14) a. Which woman did they call up?

b. *Up which woman did they call?

In (13) the *wh*-word may either be in initial position or be preceded by the preposition in the verb *call on*. However, in (14) the phrasal verb *call up* does not allow the preposition to appear in sentence initial position.

1.5.1.2 How to differentiate PVs from verbs plus prepositional phrases

According to de Haan (1988), there is also a difference between a PV and a verb plus a PP. The following example is given:

15) a. [we] [disposed] [of the problem]

b. [we] [disposed of] [the problem]

c. the problem was disposed of

d. *of the problem was disposed

Examples (15a) and (15b) above show two possible syntactic structures. In (15a), the phrase *of the problem* is a prepositional phrase, whereas in (15b) the PV *disposed of* is separated from its complement, *the problem*. Example (15c) shows that the verb plus preposition *disposed of* is a mobile PV which can be passivized; furthermore, the verb and preposition cannot be separated as in (15d). Example (15) demonstrates that *of the problem* is not a prepositional phrase in this case; in fact, *disposed of* is a PV and *the problem* is its complement.

1.6 Prepositional verbs in the literature

1.6.1 Prepositional verbs and language acquisition

One reason why it is important to investigate PVs and their use by language learners is because they are commonly found across registers, including academic prose (unlike phrasal verbs, which are most common in conversational discourse) (Biber et al., 1999). On the other hand, given their complex behavior, PVs constitute one aspect of the English language that can be challenging for many English

language learners. Tetreault & Chodorow (2008) assert that one of the reasons for such difficulty has to do with the use of prepositions, a set of forms that many—if not most—language learners find challenging (see also Hong, Rahim, Hua, and Salehuddin, 2011). This is not terribly surprising, as the researchers also report that even native speakers of English have trouble with selecting the correct prepositions. In a cloze task, only 75% of prepositions chosen were correct.

Another reason for L2 learners' difficulty in the acquisition of preposition-related collocations is that L2 students are taught and encouraged to look up unknown words in a dictionary, but are not generally encouraged to study the unknown words in context of their collocates (Hong et al., 2011). This results in unknown words being used with the inappropriate preposition, which may not necessarily affect meaning. However, it does affect the native-like quality of the learners' interlanguage.¹

1.6.2 Prepositional verbs in corpora

One way to gain insight into L2 learners' knowledge and understanding of prepositional verbs is to study how they are used in naturally-occurring language, for example, using corpus data. It has already been discussed that PVs can be found in both spoken and written texts. Furthermore, Biber et al. (1999), in their *Longman Grammar of Spoken and Written English*, report that PVs are three to four times more common than phrasal verbs, and also that PVs are relatively common in academic prose. The domains most commonly found in academic prose fall into the causative (such as *call for*) and existence (such as *refer to*, *live with*, *rely on*, *stand for*) domains. Two other common domains are activity PVs, such as *deal with*, *go through*, and *engage in*, and mental PVs such as *apply for*, *look after*, and *point to* (p. 414,418). As far as the structure of these PVs is concerned, Biber et al. also state that PVs in academic writing tend to appear in sentences with structural pattern 2 described above (NP + **verb** + NP + preposition + NP), and that these tend to be formulated in the passive voice. The table below

¹ I have seen these mistakes by learners in both conversation and writing.

illustrates some of the most common PVs used in the registers of English included in the *Longman Grammar*; namely, academic (ACAD), fiction (FICT), news (NEWS), and conversation (CONV).

Table 1.1 Prepositional verbs according to register.

| Register | Verb |
|----------|---|
| CONV | go for, go through, pay for, play with, do NP for, talk to, be expressed in, think about, listen to, worry about, know about, look like |
| NEWS | play for, serve as, be aimed at, send NP to, be accused of, be charged with, be jailed for, write to, call for, come from, be involved in |
| FICT | look at, look for, wait for, get into, turn to, stare at, glance at, smile at, be made of, fill NP with, talk about, speak to, ask for, think of, hear of, believe in, occur to, happen to, belong to |
| ACAD | deal with, be applied to, be used in, give NP to, be derived from, be divided into, obtain NP from, use NP as, refer to, add to, be known as, be seen in, be regarded as, be seen as, be considered as, be defined as, lead to, result in, contribute to, allow for, be required for, occur in, depend on, account for, consist of, differ from, be based on, be associated with, be related to, be included in, be composed of |

1.6.3 Errors in preposition selection

The literature on prepositional verbs, and specifically on error analysis, is scarce. However, Hong et al. (2011) in their article about collocations in Malaysian English learners' writing report and discuss common types of PV errors. In their study, the authors used the Oxford Collocations Dictionary and the BNC to determine acceptable collocations with prepositions. After analyzing the corpus from their Malaysian L2 English students (written and spoken data of 35,931 words), they found the most common errors of verb-noun collocations in general and categorized them by frequency into Table 1.2.

Table 1.2 Different types of errors in verb-noun collocations by frequency.

| | | |
|---|-------------|--|
| 1 | Verb | Wrong choice of verb (or non-existent verb) |
| 2 | Noun | Wrong choice of noun (or non-existent noun) |
| 3 | Usage 1 | Combination exists but is not used correctly |
| 4 | Usage 2 | Combination does not exist and cannot be corrected by exchanging single elements |
| 5 | Preposition | Preposition of a PV missing, present though unacceptable, or wrong |
| 6 | Determiner | Article missing, present though unacceptable, or wrong |
| 7 | Number | Noun used in the singular instead of the plural or vice versa |

Regarding the findings from the errors present in Table 1.2, the authors found that nearly 42 percent of collocational errors from Nos. 1 through 7 above involved the use of prepositions. Hong et al. (2011) assert that these are systematic errors which stem from overuse of the same prepositions. Out of the prepositional errors (41.72% of all collocational errors), 4.3% of errors included no preposition and in the remaining 37% contained an unacceptable or wrong preposition. Tetreault and Chodorow (2008) agree that learners overused the same prepositions. Cobb (2003) confirms this in his replication of de Cock's 1998 study; the findings showed that L2 writers were depending more on fewer MWIs. Laufer and Waldman (2011) concur, stating that L2 learners overuse a small number of learned verb-preposition collocations formed from common verbs such as *be*, *have* and *make* (p. 652).

Tetreault and Chodorow (2008) posit that there are two main types of collocation errors in relation to PVs: the selection of an incorrect preposition, and the error of omission, that is, cases in which the required "preposition is missing." In their study, they extracted data about prepositions from the MetaMetrics Lexile corpus. They found that the most common prepositions used incorrectly in general were *in*, *to*, and *of*, which are in the top ten list of most frequent prepositions.

1.6.3.1 Reasons for errors

Hong et al. (2011) provide some common reasons for errors in the use of prepositions. One possible source of error is interlingual transfer in collocations, which occurs between the learners' L1 and L2. Laufer and Waldman (2011) agree that there is an influence of L1 on L2, and cite Nesselhauf's (2005) finding that about half of collocation errors show L1 interference. Another source of PV error is intralingual transfer, which occurs within the L2 and accounts for nearly 67 percent of all collocation errors. An example of an intralingual transfer error would be **dropped into the river* instead of *fell into the river*. The largest subcategory of error within intralingual transfer is that of ignorance of rule restrictions, and an example would be **go for fishing* rather than *go fishing*. The authors state that this happens when learners extend a particular structure to a new or similar structure without considering possible restrictions.

1.7 Previous methods of analysis

Laufer and Waldman (2011) mention that the three most common ways to study collocations are through traditional analysis of errors in samples, eliciting collocations by certain techniques, and using learner corpora and associated techniques. Hong et al (2011) used the computer program WordSmith to generate the verb-noun collocations, omitting the verbs that only occurred once in order to make the list more manageable. With the remaining verbs, they generated concordance lines, which Biber, Conrad, and Reppen (1998) define as a line of text in which the target word is located between context on the left and right. Then, as mentioned before, Hong et al. used the Oxford Collocations Dictionary and BNC to find acceptable collocations so that they were consistent with previous work by Nesselhauf (2005). Any collocation which occurred less than five times in the BNC was considered incorrect and then the collocations could be classified. There were no examples given of these collocations.

To summarize, there is a great need for more research into the field of learner corpora in order to raise discussion concerning the linguistic patterns of language use within L2 interlanguage. It is apparent

that the study of collocations and specifically multi-word items (MWIs) in natural language is growing, but not yet complete. PVs are more common in written registers than their phrasal verb counterparts (Biber et al., 1999). Along with speaking, reading, and listening, writing is an important skill for all L2 learners. Therefore, it is imperative to gain understanding of how these verbs are used in natural written learner language. With these previous studies in mind, this project seeks to contribute to research and pedagogy in relation to PVs in L2 writing. Using a “bottom-up” methodological approach, this study also hopes to contribute to the current literature by helping to shorten the gap between the theory of second language learning and the interlanguage that students actually produce.

Chapter 2: Methodology

The aim of this explorative study is to obtain empirical evidence of the use of PVs in second language learner academic writing. The PVs are investigated in relation to course-based essays produced by Spanish-speaking learners of English. Using the UTEP Learner Corpus of Academic English (ULCAE), PVs are extracted using concordancing software and then analyzed for their frequency and accuracy.

2.1 Research questions

1. How frequent are PVs in the L2 writing produced by ESOL students at UTEP, as reflected in a sub-section of the ULCAE Corpus?
2. What are the most frequent PVs used by students across levels of proficiency included in the ULCAE?
3. To what extents do student essays include PVs found to be common in academic prose?
4. To what extent are PVs found in the data well formed, i.e., verbs paired with appropriate preposition?
5. What PVs, if any, seem to be most problematic for ESOL students?

2.2 UTEP Learner Corpus of Academic English (ULCAE)

2.2.1 Description

The ULCAE is a local learner corpus of general academic English comprising approximately 1.5 million words. The corpus, compiled over five years, includes first and second drafts of papers written in four ESOL classes (1610, 1309, 1311, and 1312) offered by the Department of Languages and Linguistics at UTEP, along with transcribed final writing exams given at the end of each semester. The ULCAE contains cross-sectional, quasi-longitudinal, and longitudinal data. Cross-sectional information from this corpus can be examined in order to view a sample of language from one specific period in

time, e.g., essays from Spring 2011. In addition, data from homogenous groups at different –but sequential– levels of proficiency (quasi-longitudinal data) can be examined to compare language use by students at various stages of development. Finally, the design of the ULCAE also allows for examination of the writing development of specific students across time (longitudinal data).

The vast majority of students who have contributed to the ULCAE are first language Spanish speakers who have graduated from high school in Mexico. Data are collected by written permission of students and all identifying information is removed when essays are included in the corpus and each student is then assigned a unique number. The papers are categorized by level (i.e. ESOL 1311), type of essay (i.e. evaluation essay, proposal essay, etc.), and whether it is a first, second, or only draft.

2.2.2 Description of data selected from ULCAE

A subset of essays (henceforth referred to as Sub-Corpus A) was extracted from the ULCAE for the present study, corresponding to one academic year (Fall 2011 and Spring 2012). These data include essays from the four courses/levels of proficiency in the ESOL program represented in the corpus (i.e., essays from ESOL 1610, ESOL 1309, ESOL 1311, and ESOL 1312). Sub-Corpus A includes four different types of papers per course: three different course papers composed within the context of regular course instruction and which reflect a process of revision, teacher feedback, and editing; and a final exam essay, in which no revision, editing, or feedback were available to students.

2.2.2.1 Description of writing assignments in ESOL 1610 and 1309

ESOL 1610 is an intermediate pre-degree credit English course that corresponds to level 2 in the ESOL program (low-intermediate). This course meets for six hours per week (hence 1610) and is an integrated course; grammar, reading, and writing all have a focus in this class. Students begin by writing paragraphs and then move to writing full essays. All essays generally include five paragraphs: an introduction with thesis statement, three supporting paragraphs, and a conclusion paragraph. The types of essays that are written are a description essay; a process essay, in which students describe how to do

something; a classification essay, in which students classify or divide an idea into categories that share characteristics; and finally, a final exam essay in which students chose a prompt reflecting any of the types of essays mentioned above. These essays are generally described in each course's textbook and students have the option to choose from a list of suggested topics. In ESOL 1610, writing objectives aim to help students to learn how to organize their ideas and begin writing in an academic style.

If a student is placed into ESOL 1610 and completes it satisfactorily, he or she will advance to the next level, which includes two courses: ESOL 1406, a grammar course, and ESOL 1309, the first writing-intensive course in the sequence. The latter meets for three hours per week, and it is an ESOL student's first real introduction to college writing. In this course, students write essays using "basic rhetorical patterns," as cited in the class syllabus, including cause and effect essays, in which students identify a situation and possible causes and effects; compare and contrast essays, in which similarities and differences between two items are identified and discussed; process essays, which students describe a process in detail; and finally, the students produce an essay on the writing final exam that could be any one of the types listed above. The major difference between the writing in this course and ESOL 1610 is that students may use and cite external sources for their essays, and that the instruction provided emphasizes a number of academic writing conventions, such as the use of supporting evidence and of transition words and phrases commonly found in academic prose.

2.2.2.2 Description of writing assignments in ESOL 1311 and 1312

ESOL 1311 is a composition course equivalent to the first semester of freshman English at UTEP. It is an intensive writing course based on tasks such as informing, interpreting, analyzing, and evaluating, as outlined in the class syllabus. Students use the English proficiency they gained through previous ESOL courses and build upon it while becoming more proficient in conventional academic English. Most assignments in this course focus on problem-solving for a given issue or phenomenon. The essays included in this course are: a rhetorical analysis essay, in which students choose the webpage

of a non-profit organization and examine the problems or issues such organization tries to address; an evaluation essay, in which students choose a problem, collect background information from various perspectives, analyze causal factors, and then argue one position on the topic; a proposal essay, in which students define a specific problem, seek to provide a solution based on library research and other sources of information; and finally, students write a final exam based on prompts that could pertain to the latter two types above.

ESOL 1312 is the composition course equivalent to the second semester of the required freshman English composition sequence. This course aims to build on previous English writing skills and processes while introducing the student to research-based writing. The major assignments in this course include a genre analysis, in which students choose two works about the same topic and compare and contrast them; a literature review, in which students collect and summarize background information about a topic using secondary sources; a research paper, in which students use the literature review as a basis for developing, analyzing, and synthesizing information about a specific topic; and finally, a final exam essay is written based on given prompts reflecting the type of writing produced in this course.

2.3 Method of analysis

The present study makes use of corpus-based techniques to analyze the use of PVs in essays written by ESOL students and included in the ULCAE. The first step in the process consisted of generating lists of PVs to be used as target forms. The first list came from the *Longman Grammar of Spoken and Written English* (Biber et al., 1999), which contains the 73 most common PVs. These verbs are classified according to their frequency by register: academic (ACAD), fiction (FICT), news (NEWS), or conversational (CONV) registers (see Table 1.1). In addition to the list from Biber et al (1999), a second list of PVs was compiled from the Sub-Corpus A itself by means of an analysis of prepositions (Baldwin, 2005) and their verb collocates. This second list was key to the analysis as it helped to identify any PVs that the students used and that were not listed in the *Longman Grammar*.

Finally, additional verbs were included from simple observation during transcription. The total numbers of types compiled from all three lists totaled 210 PVs. A frequency analysis of PVs was then conducted, with all target forms searched using MonoConc Pro, an automated text retrieval program.

MonoConc Pro (Barlow, 2004) is a concordancing software that is especially helpful in performing qualitative and quantitative analyses of a number of different linguistic structures. This computer program can be used to generate frequency lists of all words in a corpus, search for target words or strings of words, examine target word(s) in concordancing lines (key word in context, i.e., KWICs), locate target word(s) in their source texts, and provide information regarding target word(s) collocates to the right and left of a particular word. O’Keeffe et al. (2007) describe that concordancing lines (see Figure 2.1) enable researchers to examine instances of a word, both vertically (different forms of the word) and horizontally (other words that collocate with the target word). This allows for analysis of lexico-grammatical patterns of use of a particular language structure. For example, multi-word items collocate with each other, or are located with each other in the text. Therefore, in a written learner corpus, the target MWI can be examined and patterns can be inferred and generalized.

Figure 2.1 Sample extraction of concordancing lines.

... pace necessary to store your CDs will depend on how many CDs you own, if you have ju .
 ... e music you can have on an MP3 player depends on its memory size, if you want to have ...
 ... the variety of music you can have also depends on how much songs you buy; but it can ...
 ... he quantity of music that you can have depends on its memory size without affecting ...
 ... st important natural source, and we all depend on it. It is a fact that everyday we ...
 ... e device can hold up to 10,000 songs (depending on the capacity of player) wich means t ...
 ... ore a lot more than a thousand songs, depending in what I-Pod you have. In a Cd you can ...
 ... price of the CDs are from .25c-1dl, it depends of the capacity, for example, one disc ...
 .. are more expensive than CDs. The price depends of the capacity, for example, is more ...

Figure 2.1 provides an example of what concordancing lines look like in the program MonoConc Pro. The target word, all inflections of the verb *depend*, is located in the center with its context on either side.

Using MonoConc Pro, an alphabetical list of all words was generated for each type of essay in each level in the data analyzed. This word list was matched against the 210 PVs already selected as target forms. Matches were then searched to identify actual PVs and generate frequencies of occurrence. When searching target PVs, all inflections of each verb were included in the searches so as to not omit any PV that was present. After the PVs were identified, they were organized into a spreadsheet by register (ACAD, FICT, NEWS, CONV) and verb inflection (base form, -s, -ed, -ing, -en, and irregular forms). As an example, the following forms of the verb *think* were searched: *think*, *thinks*, *thinking*, *thought*, and *thinked*. Regarding the last in the list, it is possible that students have not yet mastered irregular past tense forms and therefore attached regular past tense affixes to the base form of the verb. Then, after data from all essays were included into the spreadsheet, the total types and tokens of PVs were calculated as well as normed frequencies per 1000 words. The purpose of norming raw frequencies from the data is to provide an equal basis for comparison between texts, regardless of length or number of words (Biber et al., 1998)

Both raw and normed frequencies were utilized to observe patterns of PV usage across levels and within levels. The normed frequencies were used to analyze the frequency of occurrence of PVs across levels (courses). Raw data were used to determine the percentage of PVs that were formed appropriately (i.e., verb + correct preposition) and also to produce a list of the top ten most frequent PVs in each level. Then, data from each of the top ten frequent PVs were analyzed for correct and incorrect usage of prepositions. Finally, data from CONV and ACAD PVs in each level were used to contrast usage within each level.

2.4 Data – Sub-Corpus A

2.4.1 ESOL 1610 and ESOL 1309

For this analysis, data from only one academic year in the ULCAE corpus was used (Sub-Corpus A). As shown in Table 2.1, the academic year of 2011-2012 includes 97,219 words collected in 201

essays from students in ESOL 1610. Generally speaking, the number of words in the description, process, and classification essays are comparable, as are the number of essays per type, although there are almost twice the number of final exams essays. The difference in number indicates that there were more final exams available to be included in the ULCAE. For this level, the average number of words per essay is 483 words.

Table 2.1 Basic information for ESOL 1610 and 1309.

| AY 2011-12 | | | | | |
|-------------------|------------|-------------|------------------|------------|-------------|
| ESOL 1610 | | | ESOL 1309 | | |
| Type of essay | # of words | # of essays | Type of essay | # of words | # of essays |
| Basic | 21615 | 41 | Cause/Effect | 16330 | 32 |
| Process | 20450 | 40 | Compare/Contrast | 18394 | 35 |
| Classification | 20139 | 42 | Process | 29724 | 48 |
| Final Exam | 35015 | 78 | Final Exam | 11607 | 22 |
| Total | 97219 | 201 | Total | 76055 | 137 |

The data from ESOL 1309 comprises a total of 76,055 words in 137 essays. The numbers for the cause/effect and compare/contrast essays are similar (32 and 35, respectively), while the process essay category contains more essays (48). In this level, the final exam essays total 22. This indicates that fewer final exams were available for transcription and inclusion into the corpus. Overall, students used an average of 555 words per essay.

2.4.2 ESOL 1311 and ESOL 1312

As shown in Table 2.2, there are 90 essays that consist of 63,591 words in the academic year 2011-2012 in ESOL 1309. The average number of words per essays is 3,706, considerably more than the averages of ESOL 1610 and 1309, 483 and 555 respectively. The higher average of words per essay reflects increased amount of research writing required. It should be noted that ESOL 1311 had far less final exams available in the corpus than other types (6).

Table 2.2 Basic information about ESOL 1311 and 1312.

| AY 2011-12 | | | | | |
|---------------------|------------|-------------|-------------------|------------|-------------|
| ESOL 1311 | | | ESOL 1312 | | |
| Type of essay | # of words | # of essays | Type of essay | # of words | # of essays |
| Evaluation | 25731 | 36 | Genre Analysis | 72171 | 74 |
| Proposal | 21154 | 27 | Literature Review | 84751 | 62 |
| Rhetorical Analysis | 13664 | 21 | Research Paper | 120568 | 62 |
| Final Exam | 3042 | 6 | Final Exam | 11415 | 19 |
| Total | 63591 | 90 | Total | 288905 | 217 |

The part of Sub-Corpus A comprised of ESOL 1312 papers is the largest with 288,905 total words. This is explained by the fact that students in this level are writing longer research papers according to academic writing conventions. This number of words includes integral parts of a research paper such as titles, research questions, headings, and references. The average number of words per essay is 1,331.

Chapter 3: Results

3.1 Frequency of PVs

As mentioned before, the first step in the procedure consisted on determining frequencies of occurrence of all target forms. The findings from this analysis are included in Table 3.1.

Table 3.1 Sub-Corpus A with normed frequencies.

| Sub-Corpus A; AY 2011-2012 | | | | | |
|----------------------------|---------------------|---------------|------------------|------------------------|---------------------------------------|
| | Type of essay/paper | # of words | Number of essays | Number of total tokens | Normed frequency of PVs per 1,000 wds |
| ESOL 1610 | Basic | 21615 | 41 | 47 | 2.17 |
| | Process | 20450 | 40 | 73 | 3.57 |
| | Classification | 20139 | 42 | 48 | 2.38 |
| | Final Exam | 35015 | 78 | 214 | 6.11 |
| | Total | 97219 | 201 | 382 | 3.93 |
| ESOL 1309 | Cause/Effect | 16330 | 32 | 113 | 6.92 |
| | Compare/Contrast | 18394 | 35 | 36 | 1.96 |
| | Process | 29724 | 48 | 86 | 2.89 |
| | Final Exam | 11607 | 22 | 100 | 8.62 |
| | Total | 76055 | 137 | 335 | 4.40 |
| ESOL 1311 | Evaluation | 25731 | 36 | 230 | 8.94 |
| | Proposal | 21154 | 27 | 86 | 4.07 |
| | Rhetorical Analysis | 13664 | 21 | 136 | 9.95 |
| | Final Exam | 3042 | 6 | 42 | 13.81 |
| | Total | 63591 | 90 | 494 | 7.77 |
| ESOL 1312 | Genre Analysis | 72171 | 74 | 625 | 8.66 |
| | Lit Review | 84751 | 62 | 575 | 6.78 |
| | Research Paper | 120568 | 62 | 888 | 7.37 |
| | Final Exam | 11415 | 19 | 103 | 9.02 |
| | Total | 288905 | 217 | 2191 | 7.58 |
| | Grand Total | 525770 | 645 | 3402 | 6.47 |

As can be observed in the cross-sectional data included in Table 3.1., ESOL students use PVs in every course/level and in every type of essay or paper included in the ULCAE. The normed frequencies included in the right-hand column clearly indicate that students increase the use of PVs as they gain proficiency in English (as determined by ESOL course/level). The frequency of occurrence of PVs in

ESOL 1610 (level 2 in the ESOL sequence) is 3.93 PVs per 1,000 words and increases to 7.77 and 7.58 PVs per 1,000 words in ESOL 1311 and 1312, respectively. One possible reason for this increase is that prepositional verbs are more commonly found in written registers, and as students become more familiar with academic writing conventions, they begin to use more relevant linguistic items, such as prepositional verbs.

3.2 Contrasting registers of PV usage

In addition to comparisons of the frequency of PVs across essay types and courses/levels, a comparison of PVs associated to different registers was conducted. For this part of the analysis, most common PVs in each one of the four registers included in the *Longman Grammar of Spoken and Written English* (Biber et al., 1999) were used, with PVs in Sub-Corpus A categorized reflecting such information. In other words, each PV type (not tokens) was classified as a PV associated to academic, fiction, news or conversational discourse. Table 3.2 summarizes these findings.

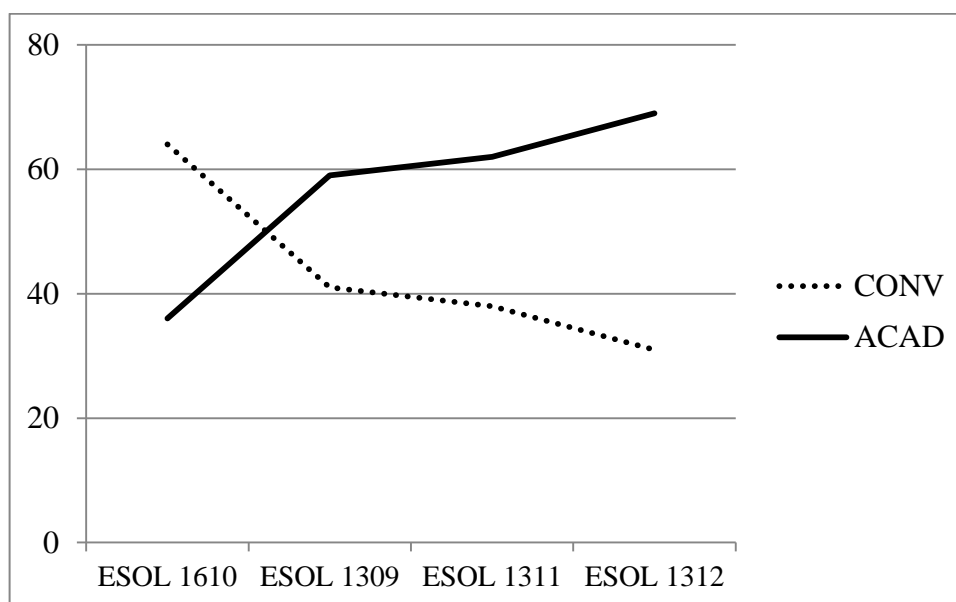
Table 3.2 Types of PVs according to register.

| Course | Type of essay | ACAD | FICT | NEWS | CONV |
|------------------|---------------------|-----------|-----------|-----------|-----------|
| ESOL 1610 | Basic | 4 | 3 | 0 | 3 |
| | Process | 1 | 3 | 1 | 4 |
| | Classification | 1 | 2 | 0 | 2 |
| | Final Exam | 3 | 5 | 0 | 7 |
| | Total | 9 | 13 | 1 | 16 |
| ESOL 1309 | Cause/Effect | 9 | 7 | 1 | 5 |
| | Compare/Contrast | 2 | 1 | 0 | 0 |
| | Process | 5 | 1 | 0 | 6 |
| | Final Exam | 3 | 5 | 2 | 2 |
| | Total | 19 | 14 | 3 | 13 |
| ESOL 1311 | Evaluation | 10 | 6 | 1 | 6 |
| | Proposal | 6 | 1 | 3 | 4 |
| | Rhetorical Analysis | 7 | 3 | 2 | 4 |
| | Final Exam | 3 | 3 | 1 | 2 |
| | Total | 26 | 13 | 7 | 16 |
| ESOL 1312 | Genre Analysis | 15 | 10 | 6 | 7 |
| | Lit Review | 22 | 11 | 2 | 7 |
| | Research Paper | 13 | 7 | 4 | 7 |
| | Final Exam | 7 | 2 | 1 | 4 |
| | Total | 57 | 30 | 13 | 25 |

Table 3.2 shows the types of prepositional verbs used according to register in each type of essay across levels. In general, ESOL 1610 and 1309 students use more CONV PVs than ACAD PVs, and as student progress through to the ESOL sequence, they use fewer CONV PVs and more ACAD PVs.

Figure 3.1 depicts the ratio of CONV to ACAD PVs in each course.

Figure 3.1 Contrast of CONV and ACAD PVs.



As shown in Figure 3.1, the usage of CONV and ACAD PVs diverge as students progress in the ESOL sequence of courses. In ESOL 1610, 64% of the total types of CONV and ACAD PVs used are CONV, and that percentage decreases as students approach ESOL 1312. On the other hand, ACAD PV usage starts off at 36% of the total of CONV and ACAD PVs and increases in ESOL 1312 to 69%. As students gain proficiency throughout the levels of ESOL, they are using fewer CONV PVs and more ACAD PVs.

3.2.1 Examples of CONV PVs in ESOL 1610 and ESOL 1309

Excerpts (a) and (b) are demonstrative of CONV PVs found in ESOL 1610 and ESOL 1309 essays.

(a) *know about*

“There are many children that already have a Facebook page, and most of them do not *know about* the privacy options.” (1610-1-11-0352-6)

(b) *listen to*

“The music is part of our lives, we *listen to* it wherever we go.” (1309-1-11-0613-1.1)

3.2.2 Examples of ACAD PVs in ESOL 1311 and ESOL 1312

Excerpts (c) and (d) show examples in context of ACAD PVs found in ESOL 1311 and ESOL 1312 writing.

(c) *lead to*

“This community development model is ecological and will *lead to* full community involvement and self-sufficiency.” (1311-1-11-0504-5.1)

(d) *be based on*

“In order to find out this information, the methodology was based on following two questions...” (1312-1-11-0345-4.2)

3.3 Most frequent PVs per course/level

The next part of the analysis focuses on the identification of the most common PVs in the data examined. Table 3.3 shows these PVs in each course/level.

Table 3.3 Most frequent PVs per course/level.

| | ESOL 1610 | Total tokens | ESOL 1309 | Total tokens | ESOL 1311 | Total tokens | ESOL 1312 | Total tokens |
|----|---------------|--------------|---------------|--------------|----------------|--------------|----------------|--------------|
| 1 | depend on* | 45 | depend on* | 41 | agree with | 90 | be related to* | 118 |
| 2 | look for | 32 | listen to** | 38 | think about** | 31 | talk about | 108 |
| 3 | spend NP on | 29 | take NP of | 31 | give NP to* | 26 | lead to* | 106 |
| 4 | pay for** | 27 | think about** | 12 | be involved in | 23 | suffer from | 106 |
| 5 | talk about | 24 | start with | 9 | talk about | 23 | focus on | 79 |
| 6 | take NP of | 22 | look for | 8 | take NP of | 22 | be based on* | 63 |
| 7 | know about** | 16 | give NP to* | 8 | know about** | 22 | be caused by | 6 |
| 8 | think in | 16 | be based on* | 7 | lead to* | 21 | take NP of | 61 |
| 9 | think about** | 15 | be made by | 7 | look for | 19 | think about** | 61 |
| 10 | give NP to* | 10 | be caused by | 7 | depend on* | 17 | give NP to* | 60 |

Note: * indicates ACAD PV; ** indicates CONV PV

Table 3.3 displays the ten most frequent PVs used in all four levels of ESOL. In ESOL 1610, students were using only two ACAD PVs (*depend on*, *give NP to*). By the time they progressed to ESOL 1312, there were four ACAD PVs present in the top ten most frequent PVs: *be related to*, *lead to*, *be based on*, and, *give NP to*. It should be pointed out that in ESOL 1311, there is quite a gap between 90 tokens of *agree with* and the second most frequent verb *think about*, which has 31 tokens. The reason for this is that the prompt for a particular essay is, “Do you agree or disagree with.....?”; therefore, students tend to use the verb given to them when responding to the prompt.

3.3.1 Examples of top two most frequent PVs in ESOL 1610

Below are two examples, (a) and (b), of the top most frequent PVs used in ESOL 1610 writing.

(a) *depend on*

“Approximately in Africa there are close to 2000 languages it *depends on* the region.” (1610-1-11-0548-3.2)

(b) *look for*

“...and the break that you are *looking for* is a good vacation...” (1610-1-11-0538-3.2)

3.3.2 Examples of top two most frequent PVs in ESOL 1309

Excerpts (c) and (d) provide examples in context of the two most frequent PVs found in ESOL 1309 essays.

(c) *depend on*

“Sometimes it changes the taste a little bit but it *depends on* what the person likes.” (1309-1-11-0508-3.1)

(d) *listen to*

“A best friend is that one who you can trust no matter what, the one who *listens to* you at any time.” (1309-1-11-0514-2.2)

3.3.3 Examples of top two most frequent PVs in ESOL 1311

Excerpts (e) and (f) demonstrate in context the top two PVs used most frequently in ESOL 1311 essays.

(e) *agree with*

“I *agree with* the fact that stereotypes are sometimes good and some other bad because...”
(1311-1-11-0584-4.1)

(f) *think about*

“We also need to *think about* how we can put into practice this recommendation in a future.”

(1311-1-11-0504-5.1)

3.3.4 Examples of top two most frequent PVs in ESOL 1312

Excerpts (g) and (h) provide examples of the top most frequent PVs used in essays in ESOL 1312.

(g) *be related to*

“Often, eating disorders *are related to* issues of co-dependency...” (1312-1-11-0364-1.2)

(h) *talk about*

“In addition, the poster *talks about* recycling processes and materials...” (1312-1-11-0453-1.2)

3.4 Correct and incorrect usage of PVs

The last part of the analysis was conducted to describe how accurately PVs are being used by students across the levels. Table 3.4 shows these findings.

Table 3.4 Number and percentage of tokens of PVs used with appropriate prepositions.

| | Types | Correct tokens | Incorrect tokens | Total tokens | % correct |
|------------------|-------|----------------|------------------|--------------|------------|
| ESOL 1610 | 51 | 278 | 104 | 382 | 73% |
| ESOL 1309 | 74 | 262 | 73 | 335 | 78% |
| ESOL 1311 | 65 | 455 | 39 | 494 | 92% |
| ESOL 1312 | 137 | 2037 | 154 | 2191 | 93% |
| | | 3132 | 370 | 3502 | 89% |

As shown in Table 3.4, at the low-intermediate level, ESOL 1610, 51 different PVs were identified out of the total set of 210 included in the list of target forms used for analysis, and students used them with 73% accuracy. Both of those numbers, PV types and percentage correct, increased in ESOL 1309, to 74 types used correctly 78% of the time. ESOL 1311 students use 65 distinct PVs appropriately 92% of the time, and ESOL 1312 experiences a jump to 137 separate PVs used 93% of the time. These data suggest that as students gain proficiency and their set of PVs grows, accuracy also increases.

Table 3.5.1 describes the percentages of correct and incorrect PVs used in ESOL 1610 and ESOL 1309.

Table 3.5.1 Correct and incorrect usage of most frequent PVs.

| ESOL 1610 | | | | ESOL 1309 | | | |
|-------------|--------|---------|-----------|--------------|--------|---------|-----------|
| PV | tokens | correct | incorrect | PV | tokens | correct | incorrect |
| depend on | 45 | 60% | 40% | depend on | 41 | 73% | 27% |
| look for | 32 | 66% | 34% | listen to | 38 | 42% | 58% |
| spend NP on | 29 | 34% | 66% | take NP of | 31 | 90% | 10% |
| pay for | 27 | 41% | 59% | think about | 12 | 100% | 0% |
| talk about | 24 | 92% | 8% | start with | 9 | 100% | 0% |
| take NP of | 22 | 59% | 41% | look for | 8 | 100% | 0% |
| know about | 16 | 100% | 0% | give NP to | 8 | 100% | 0% |
| think in | 16 | 100% | 0% | be based on | 7 | 71% | 29% |
| think about | 15 | 100% | 0% | be made by | 7 | 100% | 0% |
| give NP to | 10 | 100% | 0% | be caused by | 7 | 86% | 14% |

ESOL 1610 students used a mix of registers of prepositional verbs, including ACAD and CONV.

Depend on is the most common PV used, but they used it correctly 6 out of 10 times, as seen in Table 3.5.1 above. However, they used CONV verbs such as *talk about*, *know about*, and *think about* with near perfect accuracy. ESOL 1309 students used *depend on* the most frequently as well, but accuracy improved to 3 out of 4 times correct. They also were extremely accurate with less academic PVs such as *think about*, *start with*, and *look for*.

3.4.1 Examples of incorrect usage of PVs in ESOL 1610 and 1309

The examples (a), (b), (c), and (d) below outline some instances of incorrect usage of PVs in the lower level courses, ESOL 1610 and ESOL 1309.

(a) *spend NP on*

“Try to *spend less money in* expensive devices, clothes or other things...” (1610-1-11-0541-6)

(b) *pay for*

“...because students have to *pay* many things like books, food, transportation, etc.” (1610-1-11-0542-6)

(c) *be based on*

“The buildings in each city *are based in* the conditions of the zone.” (1309-1-11-0618-2.1)

(d) *listen to*

“...they will turn on the radio and *listen* music and relax with it” (1309-1-11-0613-1.1).

Example (a) was extracted from an essay in ESOL 1610. There is a preposition present (*in*); however, it is incorrect. In example (b), the excerpt is also from an ESOL 1610 essay. In this example, the preposition is missing. Example (c) was obtained from a portion of an essay in ESOL 1309, and in this example the preposition is incorrect (*in*). Finally, example (d) shows a missing preposition in an ESOL 1309 essay.

Continuing with the error analysis, percentages of correct and incorrect usage of the most frequent PVs in ESOL 1311 and ESOL 1312 are shown in Table 3.5.2.

Table 3.5.2 Correct and incorrect usage of most frequent PVs.

| ESOL 1311 | | | | ESOL 1312 | | | |
|----------------|--------|---------|-----------|---------------|--------|---------|-----------|
| PV | tokens | correct | incorrect | PV | tokens | correct | incorrect |
| agree with | 90 | 100% | 0% | be related to | 118 | 76% | 24% |
| think about | 31 | 84% | 16% | talk about | 108 | 100% | 0% |
| give NP to | 26 | 100% | 0% | lead to | 106 | 100% | 0% |
| be involved in | 23 | 87% | 13% | suffer from | 106 | 65% | 35% |
| talk about | 23 | 100% | 0% | focus on | 79 | 79% | 21% |
| take NP of | 22 | 100% | 0% | be based on | 63 | 93% | 7% |
| know about | 22 | 96% | 4% | be caused by | 6 | 94% | 6% |
| lead to | 21 | 91% | 9% | take NP of | 61 | 89% | 11% |
| look for | 19 | 95% | 5% | think about | 61 | 92% | 8% |
| depend on | 17 | 83% | 17% | give NP to | 60 | 100% | 0% |

ESOL 1311 students use more ACAD verbs more frequently, including *give NP to*, *lead to*, and *depend on*. The latter is not used as frequently as in ESOL 1610 and ESOL 1309, but accuracy increased to 83%, as shown in Table 3.5.2. CONV PVs were used correctly almost 100% of the time. ESOL 1312 students employed more ACAD verbs among the most frequent verbs: *be related to*, *lead to*, *be based on*, *give NP to*.

3.4.2 Examples of incorrect usage of PVs in ESOL 1311 and ESOL 1312

Below, excerpts (e), (f), (g), and (h) provide inappropriate usages of PVs in the upper level writing courses, ESOL 1311 and ESOL 1312.

(e) *think about*

“...because adults do not *think on* cyber bullying like a serious problem.” (1311-1-11-0492-4.1)

(f) *depend on*

“...both informative and persuasive text *depending in* if you are the one who needs help...”

(1311-2-12-0590-7.2)

(g) *suffer from*

“...Health Care for the Homeless argued that 39 percent of the homeless *suffer a* mental illness.”

(1312-2-2012-0489-3)

(h) *be related to*

“Also stress *is related with* physical illness such as hypertension and some headache...” (1312-2-12-0579-3)

In example (e) extracted from an ESOL 1311 essay, the target form was not met as there is an incorrect preposition (on) following the verb (think). Example (f) demonstrates a verb that has shown to be problematic in emergent academic writing. In this example, the preposition is incorrect (in). Example (f), extracted from an ESOL 1311 essay, shows that the preposition is missing. Finally, in example (g) which was isolated from an ESOL 1312 essay, the preposition is incorrect (with).

As seen in Chapter 3, both a frequency analysis and error analysis were conducted in order to determine the patterns of frequency across ESOL levels/courses, as well as to determine the patterns in appropriate and inappropriate usage of PVs.

Chapter 4: Discussion and Implications

In this project, I have explored frequencies of occurrence and levels of accuracy in the use of PVs by examining corpus data produced by L2 English students at different levels of proficiency. The aim was to contribute to our understanding of how Spanish-speaking learners of English use PVs in their course writing assignments and to outline possible pedagogical implications. As the research on this area suggests, it is imperative for English learners to acquire and master multi-word items in order to use written and spoken English in a more fluent and effective manner (Laufer & Waldman, 2011; Hong et al., 2011). However, it is also known that the acquisition of PVs is difficult for L2 learners. Studies like the one reported here can help to make more focused and adequate classroom decisions regarding type and amount of instruction and feedback provided to students in relation to multi-word items.

4.1 Discussion

In general, the learner corpus data extracted from the ULCAE provided enough information to answer the research questions outlined for this project. The findings clearly show that ESOL students use prepositional verbs at all levels of proficiency, and that the frequency increases throughout the levels, from 3.93 per 1000 words in ESOL 1610 to 7.58 per 1000 words in ESOL 1312. As students gain proficiency in their writing skills, they are using more prepositional verbs in general. A surprising finding was the overall high accuracy of student usage of PVs. In the low-intermediate English course, ESOL 1610, students used 51 different types of PVs correctly nearly 75% of the time. The number of types of PVs used and accuracy of use only increased as students moved through the levels. It must be pointed out that, unlike conversational English in which speakers do not have an opportunity to edit and revise, the subset of essays examined consisted of second or final drafts which have gone through a process of revision and editing. In the ESOL program at UTEP, students have access to feedback from writing tutors, peers (during peer review sessions), and instructors before they submit the final draft of the assignment. This could be a reason why the accuracy is so high; at the same time, it displays

students' attention to the writing process by demonstrating willingness to obtain constructive criticism to improve their writing.

In addition to feedback, it is possible that students improve their use of PVs based on increased oral input from their teachers and classmates. At the university, students of all levels of ESOL have an opportunity to attend informal conversation sessions, in which the conversation session leader chooses a topic and leads whole-class, group, and pair discussions. Each session is an environment where the students receive input from each other and from the leader. This could account for the high accuracy (> 90%) of CONV PVs that was found in writing at the intermediate level. Yet another possibility for increased proficiency and accuracy of PVs in all levels is that students have simply acquired the PVs as single units.

It was encouraging to find that students used both academic and conversational PVs even at the low-intermediate level (ESOL 1610). As explained previously, many of the 210 types explored are assigned a register in the *Longman Grammar*; these registers are ACAD, FICT, NEWS, and CONV. In the most commonly used PVs in ESOL 1610, students used only two ACAD verbs: *depend on* and *give NP to*, while they used more conversational verbs such as *talk about*, *know about*, and *think about*. It should be noted that they used the latter verbs with high accuracy (more than 90%). CONV PVs are used frequently in conversation and informal language, and these forms were more frequently used at the intermediate level in the ESOL program.

In contrast, four out of ten of the most common PVs used in ESOL 1312 are academic: *be related to*, *lead to*, *be based on*, *give NP to*, which suggest a shift in the type of PVs that students at the more advanced levels use in their writing. Students use these verbs with accuracy of three out of four times or better. In addition, out of the most common verbs used in ESOL 1312, two of the academic PVs are used successfully in passive voice. Consequently, it can be seen that students are using PVs along other conventions for writing in academic English.

Another indication that students are using fewer CONV PVs and more ACAD PVs in their writing can be inferred when comparing the usage of just those two registers of verbs. What was found was that students are in fact “trading in” their less formal verbs for PVs that are associated to more academic discourse. The reasons for this could be several. First, many of the writing prompts, or questions, given in class and on the final exams already contain ACAD PVs. For example, in ESOL 1312, there is a prompt that reads, “X *leads to* Y.” Students then use the PVs provided in their writing. Another reason could be that the curriculum contains more ACAD PVs since ESOL 1312 is a research writing intensive course. A third reason could be that students have acquired these verbs through readings and other types of academic input (lectures, workshops) and use them naturally in their interlanguage.

Despite the evident progress that ESOL students have made in the number and way PVs are used in their writing, there are still a few PVs that are problematic. One that is prominent is *depend on*. Native English speakers use this verb quite frequently, and may not think twice about the appropriate preposition. However, the data show that ESOL students, even at a more advanced levels, are not using the PV *depend on* accurately. A most likely cause in this case may be transfer from the first language (Hong et al., 2011; Laufer and Waldman, 2011). Most students are L1 Spanish speakers, therefore it can be assumed that they are transferring their knowledge of *depender de* (literally “to depend of”) to English, consequently using an incorrect preposition in English. Another example of a PV that ESOL students struggle with is *consist of*, in which students generally write *consist in*. Again, harkening back to their L1 experience, the verb in Spanish is *consistir en* (literally “to consist in”). An obvious question to answer would be providing a remedy in pedagogical terms for situations such as L1 transfer. It is

believed that in such circumstances, students would benefit most from “form-focused instruction” (Laufer and Waldman, 2011).²

4.2 Limitations

As it was mentioned before, there is a lack of information in the literature about L1 Romance language speakers and their use of PVs in speaking and writing. This study examines mostly L1 Spanish speakers’ writing, so while it does provide some very useful data, there can still be more work done across other L1 Romance language speakers of French, Portuguese, Italian, etc. However, in order to replicate this study, there would have to be appropriate learner corpora available.

Another limitation of this study is that the list of PVs used to search target forms in the corpus data selected was not exhaustive. It is possible that ESOL students use PVs not included in such a list. However, the examination of preposition-related collocates provides a certain level of confidence that most PVs used by these students were identified.

4.3 Suggestions for further study

In this particular corpus-based study, second drafts were used along with final exams; therefore, these data include writing that was both exposed to corrective feedback and writing that was not. While the results from these data were insightful, it would also be informative to extract only final exam essays to determine how ESOL students use PVs in a setting where feedback from tutors, peer review sessions, or instructors is available.

In addition to a study on how PVs are used in final exam essays, a more in-depth error analysis would shed light on exactly which errors ESOL students are making when using PVs without the benefit of corrective feedback. Hong et al. (2011) in their error analysis propose sources of errors, including L1

²This study confirms my observations of these particular verbs during my time as an ESOL writing tutor, and I aimed to give direct instruction by explaining the differences between the preposition in Spanish and how it is said in English. This approach seemed to be very helpful to students.

interference and intralingual transfer. Further study could reveal information about these possible sources of error in UTEP's local learner corpus.

The difficulty posed by prepositions for English language learners is well known. According to the study done by Baldwin (2005), there are 34 prepositions alone, and the combinations with verbs are virtually endless. How can L2 instructor help learners to acquire the forms of prepositional verbs? Research suggests that one way for learners to acquire not only prepositional verbs but all MWIs is through direct instruction (Gardner & Davies, 2007; Hong et al., 2011; Cobb, 2003). Cobb (2003), in relation to the acquisition of these forms, states that "even advanced learners are unlikely to discover very quickly on their own all of the relevant features of a second language that make it native-like" (pp. 418-419). Therefore, one possible instructional approach is to focus on those PVs proven to be problematic across levels and address these forms directly through instruction. Another alternative would be to highlight specific differences between prepositions used with certain PVs in English and their equivalents in the students' L1.

It may seem like a daunting task at first, but with information generated from empirical research in this area, including frequency and error analyses, these findings can be integrated into the classroom setting. Nesselhauf (2005) suggests that collocations should be taught "systematically and according to the degree of difficulty, from the most to the least difficult" (cited in Hong et al., 2011, p. 42), while Laufer and Waldman (2011) are of the opinion that collocations and MWIs should be taught by the following: eliciting the collocations, matching the words together, selecting the missing word, and completing collocations from memory. Learner corpus data collected locally can contribute to the instructional purposes of the local institution(s) because the results of analyses done can be applied directly to the learners who generated the data in the first place. Further work in the field of multiword items and preposition-related collocations can help English language learners become more competent and fluent in the use of the *lingua franca* of our time.

References

- Baldwin, T. (2005). Looking for prepositional verbs in corpus data. In *Proceedings of the 2005 Meeting of the Association for Computational Linguistics: Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications* (pp. 115-126).
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating structure and use*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finnegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Longman.
- Bonelli, E. T. (2012). Theoretical overview of the evolution of corpus linguistics. In A. M. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 14-27). Routledge.
- Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 59(3), 393-424.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32(1), 17-44.
- Flowerdew, L. (2012). *Corpora and language education*. Palgrave MacMillan.
- Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *Teaching English to Speakers of Other Languages Quarterly*, 41(2), 339-359.
- Granger, S. (2002). A bird’s-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.) *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 517-533). John Benjamins.
- Hong, A. L., Rahim, H. A., Tan, K. H., & Salehuddin, K. (2011). Collocations in Malaysian English learners’ writing: A corpus-based error analysis. *3L; Language, Linguistics and*

Literature, The Southeast Asian Journal of English Language Studies., 17(special issue), 31-44.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647-672.

O'Keeffe, A. M., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Quirk, R. (1972). *A grammar of contemporary English*. New York: Seminar Press.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, 10(1-4), 209-232.

Teschner, R. V., & Evans, E. E. (2000). *Analyzing the grammar of English*. Georgetown University Press.

Tetreault, J. R., & Chodorow, M. (2008, August). The ups and downs of preposition error detection in ESL writing. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume* (pp. 865-872). Association for Computational Linguistics.

Tsui, A. B. (2004). What teachers have always wanted to know—and how corpora can help. In J. H. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 39-66). John Benjamins.

Urzúa, A. (accepted for publication). Corpora, context, and language teachers: Teacher involvement in a local learner corpus project. In V. Cortes and E. Csomay (Eds.), *Corpus-based research in the twenty-first century: Developments from the School of Flagstaff*. Philadelphia, PA: John Benjamins.

Vita

Elizabeth M. (Little) Wilcoxon was born and raised in central Illinois. She was always a reader, amateur linguist and grammarian throughout school, and after high school she pursued a BA at Bradley University in Spanish and secondary education. She then began her professional teaching career at Midland High School in the small farming community of Varna, Illinois. She relocated to El Paso, Texas in 2011, and in 2012 she began graduate studies in the Languages and Linguistics department at the University of Texas at El Paso. In 2014, she won the Graduate Student Instructor Teaching Portfolio Award, was honored as an Outstanding Graduate Student in Linguistics, and also presented this research at the TESOL Master Student Forum in Portland, Oregon. She has aspirations to acquire a PhD and become a professor someday, but for now she would like to continue teaching and honing her research skills in the fields of applied linguistics and TESOL.

Permanent address: 9816 Sidewinder St
El Paso, Texas 79924

This thesis/dissertation was typed by Elizabeth M. Wilcoxon