A Lexical Error Analysis of Advanced Language Learners' Writings

By

Nathan Wells

A Thesis

Submitted in Partial Fulfillment of

the Requirements for the Degree of

Master of Teaching English to Second Language Learners

Vladimir Pavlov	Date

Annette Klemp	Date

Laurel Langford	Date

Director of Graduate Studies	Date

University of

Wisconsin-River Falls

2013

Advanced language learners have unique challenges with vocabulary instruction. At the advanced level, learners most likely have an extensive vocabulary that covers the high frequency words of the langauge. However, traditional vocabulary instruction focuses on creating breadth in vocabulary and not depth. This approach does not meet the needs of ALLs. This study aims to provide instructional suggestions through a lexical error analysis of ALL writing. The area of lexical errors in advanced language learners (ALLs) has been researched very little. This study takes a look at the lexical errors committed by advanced language learners in the university setting. The purpose of the study was to determine what types of lexical errors ALLs make, the influence of first language direct translations on lexical errors, the affect of split category cases on lexical errors, and pedagogical implications for ALL vocabulary instruction. The study uses a corpus consisting of thirty-one essays from Russian speaking and Korean speaking students at the post-secondary level. The essays were analyzed for lexical errors. Those lexical errors were then sorted and classified. Two surveys were then administered to determine the likelihood of calques and split categories as an influence on lexical errors. It was found that a statistically significant number of lexical errors were made. Over 50% lexical errors had to do with the learner not understanding the semantic range of the word and not understanding appropriate collocations of the word. In light of the findings, several approaches and activities are provided to use with ALLs. The focus of the activities are to create individualized and differentiated instruction through the use of student writing and goal setting. The activities also focus on giving ALLs a deeper knowledge of vocabulary by using semantic mapping, studying collocations, and using concordances.

Table of Contents

Introduction

When one enters the discipline of language instruction, he is faced with an abundance of theories, methods, and strategies. Experts from the fields of psychology, sociology, linguistics, education, and the cognitive sciences weigh-in on the subjects of language acquisition and the best pedagogical practices. One can view language as an isolated ideal entity with a set number of basic parameters and principles, or it can be viewed as an entity that exists as a result of the brain being able to make generalizations from a massive amount of input. The language instructor needs to consider many different theoretical views and instructional methods.

This study attempts to present one small portion of this educational paradigm. The ultimate aim of this study is to look at the types of lexical errors that are being committed by advanced language learners (ALL) in their academic writing and then provide some pedagogical suggestions to deal with the specific errors. The purpose of the first chapter is to create an understanding of the complex nature of a learner's vocabulary, different components of a lexicon, and the importance of vocabulary in language acquisition. The chapter begins with a basic explanation of what a lexicon is and its components. This is followed by how a word is defined. As one will see, the task of defining a word is no small task. Depending on the focus of the discussion or what a researcher is looking to prove or disprove, the word can be defined in many different ways. This section explores many of the different approaches currently used in modern linguistics in order to shed light on the choices made by the research in this paper. Similarly, there are many layers and gradients in knowing a word. Since this paper aims to assess the language learner's knowledge of words in her mental lexicon, it will be important to understand all these levels and gradients. Another key aspect of the current research is a learner's

understanding of a given lexical item in the context of the words surrounding it. The section on lexical phrases explores the various types of set phrases, how they are connected, and what sets them apart. Because both native speakers and second language learners utilize set phrases, the following section explores the uniqueness of collocations with second language learners and some previous research that has been done on this topic. The current research attempts to bring further clarity to the topic of collocations with second language learners. Chapter one concludes with a brief discussion on the importance of vocabulary in the acquisition of both the first and second language.

Because the approach of this research is a lexical error analysis, Chapter two reviews literature on lexical error analysis. Many of the ideas, classifications, and techniques of previous literature were a strong influence on how this research was approached. In Chapter two, the different parameters of for lexical error analysis are explored. Also included are the different ways that lexical errors have been classified. This section aims to clarify the choices made in the current research.

The purpose of Chapter three, which discuss vocabulary instruction, is to provide a framework for the pedagogical suggestion made at the end of the paper. This chapter does not solely focus on language instruction in a second language, but rather, it takes a broader approach and discusses language instruction in both the first and second language. There are a couple of reasons that this approach is taken.The first is that in current second language vocabulary instruction seems to lack in certain areas. This becomes evident as we look at the sections on lexical error analysis and set phrases. The second reason for taking a broader look at language instruction is that in order to make sound pedagogical decisions in second language vocabulary

instruction, one needs to understand the approaches taken with first language vocabulary

instruction, and that some of these approaches may not be suitable for second language

vocabulary instruction. Contrarily, after the findings of this paper and the ones preceding become

evident, one may find some first language approaches are appropriate for a second language

learner.

Chapter four is on the current research of this paper. It lays out methods used, the

definition of the *word* as it relates to this study, and the parameters for a lexical error. The chapter

also describes the subjects who participated in the study, and the different tools and procedures

that were used to analyze learners' vocabularies. Chapter four ends with the results from the

research and a discussion on what the findings point towards. The findings of this research

guided the choices made in the final chapter which discusses pedagogical suggestions for

advanced language learners.

Finally, Chapter five provided suggestions for teaching vocabulary to advanced language

learners. As seen in Chapter three and four, advanced language learners face unique issues in

regards to vocabulary instruction. It is no longer about the breadth of vocabulary instruction, but

rather its depth. Chapter five provides approaches and specific activities designed the help the

learner identify and focus on the deeper meanings of words that are relevant to her learning. The

presentation of these approaches and activities are an attempt to give readers tangible tools that

they can easily implement into their instruction.

1. The Parts and Importance of Vocabulary

This paper functions on several fundamental beliefs that make this study relevant. The first belief

is that vocabulary is one of the most important parts of language; however, the area of

vocabulary has been neglected in the research field of language instruction for the most part. Until recently, it has been hard to find any research on the influence of vocabulary in language learning. The second belief is there also seems to be a gaping hole in vocabulary instruction for language learners at different levels. As we will see in this paper, there are many suggestions for teaching vocabulary, but there are few real suggestions for the best instructional methods to be used with advanced learners. This latter belief is perhaps the most important because it is what makes this study relevant. It is the view of this paper that advanced learners face different challenges in vocabulary use than other learners, and traditional methods of vocabulary instruction do not address the needs of advanced learners. The final belief is that with the advancement in technology, the use of corpora to study language-in-use can be an effective tool to figure out best pedagogical practices.

Before an inventor can improve an existing item, she needs to understand that item. She needs to understand how the item is constructed, what its current purpose is, where it originates, and what its current shortcoming are. The same is true with creating sound pedagogical instruction and educational research. The educator needs to understand how knowledge is acquired, how it is constructed, what its parts are, how the knowledge is used, and what the current deficits are in the learner's language. Then, through research, the educator is able to make sound decisions on the activities and methods that will best suit the needs of the learners.

1.1 Lexicon

This section will move from the whole to the part and then move back towards the whole. It will look at what a lexicon is and what the different parts are. Let us briefly consider the make up of one's lexicon. *Lexicon* comes from Greek meaning "dictionary" ("lexicon"), so a person's

lexicon is their personal dictionary. As with most dictionaries the human lexicon is thought to consist of words (simple, complex, and compound). An example of a simple word is *teach*, a complex word is *teacher*, and a compound word is *doghouse.* There are also phrases like idioms (*kick the bucket*) and collocations (*knead the dough*). A word may be represented by a phonological, orthographical or imagerial symbol. If an individual knows a word she, recognizes the sound of a word, she can often spell it, and, in most cases, she can also conjure a mental image of the word. There is also information on the word's denotata, connotations, and grammatical classifications. In a first language lexicon, all these pieces work seamlessly together to allow the individual to choose the correct lexical item needed in order to communicate effectively. All of these items are naturally and effortlessly acquired by a child through natural interaction with the language.

1.2 The Word

Word is a very difficult word to define in that there is no definition that completely explains the term. There is still much debate amongst scholars about what constitutes a word, probably because the word is multi-faceted, and whichever definition is given, one is able to find an instance where the definition does not hold true. For this reason, the next portion will briefly discuss the different aspects of the word, and then the definition which was used for this study will be given.

The meaning of *word* can be defined by several different criteria. If we look at the word from the perspective of the writing tradition, a word could be defined as a sequence of uninterruptible graphemes, which is set off by spaces preceding and following it (Jackson &

Amvela 48). The obvious problem with using this perspective is that there is no thought given to meaning and *ljeet* could be a word.

We can also consider the word as a unit of thought (Jackson & Ambela 48). It can be one item, such as *book* and have one corresponding thought unit, which would be the physical object in the real world. It can also be one item and have two thought units, such as *farmer* which would be *a person* who *farms*. Finally, it can be multiple items, such as *all of a sudden* and have only one unit of thought (Jackson & Amvela 52). Once again, if the criterion of meaning is only used to define the word, then what is to be done with the examples *bad* and *awful*. Both could be used to describe something of low quality or of something undesired, but one would never consider them to be the same word; therefore, it is clear that both form and meaning must be taken into consideration.

Function or word class could be another criterion for defining a *word* (Hatch & Brown 220). The example *farm* could be either a verb or noun depending on its position in the sentence. In this example, the form is the same, but the meanings are different. As a noun, it denotes an area of land where crops are grown or animals are raised for sale, and as a verb it denotes the action of using the land for growing crops or raising animals. For apparent reasons, function or word class is not the only acceptable criterion for defining the *word*, just as using form or meaning as the only criterion would not be acceptable to define the *word*. If function were the only criterion, then *pen* and *pencil* would be the same word because they both hold the function of noun. As we will see, all three of these criteria will be used to define the current study's definition of a unique word.

 Consider the following example: *am, is, are, was,* and *were*. All five of these words have different forms, and their meanings are different when grammatical categories such as first person present singular and third person past plural are taken into consideration. However, they all function as a verb. So, the question is: Are these different words? In order to answer this question, we need to consider the idea of a grammatical items and lexical items.

One approach to the *word* is from a grammatical position. I stated above that *am, is, are, was,* and *were* have different forms and, from a certain view, meaning. They are all derived from the same verb *to be* and they fill different grammatical categories in terms of person and tense. *Am* is in the first person singular form, and *were* is in the third person plural past form. So, if one wishes, he could say that because *am, is, are, was,* and *were* have different forms and different grammatical categories, then there are five different grammatical items and five different words. The same approach could be taken with *book* and *books*. Both words have different forms, and they are in different grammatical categories, so, from a grammatical stance, they are two different words. Another approach to the word is from a lexical stand-point. From this position, the examples *am, is, are, was,* and *were* are one lexical item, or word, because they have the same meaning and are derived from the same verb *to be*. Also, *book* and *books* are one lexical item, or word, because they have the same meaning. In short, the changing of the grammatical category does not constitute a new item, however, a different function (verb, noun, adverb, adjective, etc.) does constitute a new lexical item. For example:

The *farm* has been in the family for years.

I will *farm* for the rest of my life.

He is a *farmer* on this *farm* and on two other *farms*.

In the first two examples, *farm* is functioning as a noun and as verb in the second example. Although their forms are the same, the different functions give a different meaning and therefore create two distinct lexical items. The third example is included in order to make the comparison between inflections and derivations. *Farms* has the plural inflection *–s* and as we said above *book* and *books* are one lexical item, so, *farm* and *farms* are one lexical item; in contrast, *farm* and *farmer* are two different lexical items because *farmer* has the derivation *-er* which changes the function. Therefore, since inflections simply change the grammatical category and not the meaning, they do not constitute a new lexical item, and since derivations change function and the meaning they do constitute a new lexical item.

Finally, we can consider the word in terms of *free* and *bound* forms. A minimal form is a morpheme because it is the smallest unit of language that carries meaning; therefore, a *free form* is a morpheme that can exist alone by itself and have a full lexical meaning. For example, the morpheme *play* can exist alone and meaning can be derived from it. It can either be a noun denoting a theatrical production, or it can be verb denoting the taking part in an enjoyable action. A *bound form* is a morpheme that needs to be a bound to a free form. Some examples are the morphemes *–s, -er, pre-,* and other affixes. Bloomfield, as cited by Jackson and Amvela, suggests defining a word as, " a minimal free form" (49).

A pivotal point in this paper is the idea that the 'word' can exist as a unit larger than the single item. It is the idea that a 'lexical phrase' can function in the same way as a single word; therefore, it is stored and used as a single lexical item. This includes certain word combinations like collocations and idioms.

To conclude, to define *word* is not an easy task. The criteria used for defining a word depends greatly on the reason one wants to define the term. A high school student who needs to write a 500 word essay for a final exam would most likely include a graphical and semantical criteria and not pay much attention to the number of thought units included in a word or set of words. A lexicographer building a dictionary for a beginning native language learner would most likely take into consideration graphical, semantical, grammatical, and though unit criteria. For this study, the criteria used to define a word are graphical, semantical, grammatical, and thought units. All these criteria are used to create a narrow and specific definition in order to isolate the reasons for the mistakes that ALLs make.

1.3 Knowing a Word

Before looking at the different ways of classifying the learners' errors, it is important to consider the root of these problems. In regards to the latter, one possible reason for lexical errors is related to the fact that knowing a word is not as simple as being able to produce or recognize it. There are many aspects to knowing a word.

Knowing a word is usually a dichotomous split between form and meaning (Gathercole & Baddeley 72). The form exists in two different ways. Knowing the spoken form is being able to recognize the word's phonological form when hearing it and being able to speak it (Nation 40). Following Nation's logic, knowing the written form of a word is being able to understand the word when encountered in reading and being able to use it when writing. Therefore, a person knows a word's form if he can speak it, write it, recognize it when heard, or read it. A connection can be made here to the receptive vocabulary and the productive vocabulary. A learner could have word form knowledge in both the receptive and productive vocabulary, but this does not

meant that he fully knows the word. A leaner could hear a word and recognize it, write it, read it, and speak it, but not have the meaning solidified in his lexicon. Therefore, there is more to knowing a word than knowing its form.

In regards to meaning, Robinson states, " [knowing] a word involves both knowing its denotative meaning and how it can, and is, being used" (275). The latter part of this statement is relatively obvious. A learner should know the denotative meaning of the words, but the former part brings to light another important aspect that is often over looked in vocabulary instruction. Often times, there are many words that have the same or similar denotative meanings and these are known as *synonyms*, but the definition of a *synonym* is that they are a set of words that can be interchanged in some context, but not in others. The reason that synonyms are not interchangeable in some contexts is because each word has certain connotational meanings, fine nuances that are not explained in the denotative dictionary definition. A major part of knowing a word is knowing how a word functions and operates in different contexts. There is often a difference between a words isolated meaning and phrasal meaning (Shahheidaripour 5).

For example, when looked up in a thesaurus, the words *factor*, *reason,* and *thing* appear as synonyms; they have the same denotative meaning (Thesaurus.com). Now consider using these words in a formal academic essay. *There were many contributing (factors/reasons/things) to the outcome of this study.* In the context of a playground conversation, *Why do think Tommy is so mean? I think there are a lot of (factors/reasons/things) that make Tommy mean.* In the first sentence, which of the three words is acceptable? In the second sentence, what would happen to the child who uses the word *factors?* If a leaner is to know any of these three words, she would need to know in what contexts these words are acceptable. The examples given above show us

that different words have different levels of formality. They are used in different registers. Some words have very formal aspects to them, such as *factor*. The word *factor* is more likely to be used in an academic environment. Other words are more informal and used in the colloquial contexts. The words *reason* and *things* are both examples of this. These words would most likely be used in everyday conversation. The learner must have an understanding of appropriate contexts of words in order to use them successfully.

One final consideration in knowing a word is knowing its grammatical function. Grammatical function is a words part of speech and the grammatical patterns it fits into (Nation 55). If someone truly knows the word *chair*, he knows that it can function as a noun (*He sat in the chair* or *He is the chair of the committee*) or as verb (*He is chairing the committee*). The grammatical function is often dictated by the surrounding context (Ellis 96). If a learner does not know all the possible grammatical functions, she may become confused when it is encountered in an unfamiliar context. Furthermore, the leaner may try to use the word in an unacceptable way.

To conclude, there are many aspects to knowing a word. Nation presents a list of the different aspects to knowing a word: spoken form, written form, word parts, connecting form and meaning, concepts and referents, associations, grammatical functions, collocations, and constraints on use (40-59). Knowing a word's spoken form is being able to recognize it when it is heard and being able to produce the word to communicate meaning. Knowing a word's written form mirrors knowing its spoken form. A learner should be able to recognize a word when encountered in a text and be able to accurately produce its written from in writing. Knowing a word's parts means that the learner knows the meaning of a word's root and affixes (both

inflectional and derivational). Nation states that it is possible for the learner to know just the form or just the concept. The learner needs to know both the form and the concept. Nation also suggests that it is possible that a learner knows the form and the concept, but it is not connected in the mental lexicon. To know a word's form and meaning, a learner must have both of them and have them connected. A learner needs to know the word's concepts and referents. This means that the learner needs to know the different denotational meanings of a word. This is connected to concepts of polysemy and homonymy. Nation's criterion of 'associations' is related to how different parts of speech can be organized. For example, nouns can be classified into hierarchal trees such as mammals, reptiles, etc. Adjectives can be classified into predicative and non-predicative. Also, verbs can be classified into event verbs and state verbs. Knowing these different classifications can help with the next criterion. Grammatical functions refer to word classes. In order to know a word the leaner must know what word classes and grammatical patterns the word functions in. Another part of knowing a word is being acquainted with other words that it commonly occurs with. A learner needs to know that we have *fast food* and not *speedy food* (56). Finally, in order to know word, a learner must know its usage constraints. This is similar to knowing its connatational coloring. For example, words occur in certain registers, geographical areas, with certain frequency and currency, and have certain literal and metaphorical meanings (57). As we can see, there are many different aspects to know a word, which in turn, means that there is much lexical knowledge that needs to be acquired by the learner. This leaves much room for the leaner to err.

1.4 Set Phrases

In general, the concept of a set phrases is the idea that some words have a tendency to co-occur with other words. These tendencies can range from loose relations with interchangeable pieces to strict fixed word orders with opaque meanings. As will be seen, this simple idea has some very complex undertones that have many experts shading in grey areas and mixing terms and concepts in order to define *word*. For example, at one point in the discussion, Hogwarth classifies idioms as collocations; however other researches would classify an idiom separate from a collocation under the larger term *set phrases*. The current research has a strong focus on collocations; however, in the following section, time will be given to a broader discussion of all types of set phrases.

Singleton gives the definition of a collocation as "frequently occurring combinations" of words (11). Jackson and Amvela offer a more complex definition, "Collocation refers to a structural or syntagmatic relation, to meaning relations that a word contracts with other words occurring in the same sentence or text" (113). Furthermore, Nation defines the collocation as a "group of words that belong together, either because they commonly occur together, or because the meaning of the groups is not obvious from the meaning of the parts" (317). Finally, Benson gives the following perspective, " In the English language, there are many fixed, identifiable, non-idiomatic phrases and constructions. Such groups of words are called recurrent combinations, fixed combinations, or collocations" (qtd in Bahns & Eldow 102). Benson isn't the only one who has given the collocation different labels. As stated before, the concept of set phrases is a very complex subject with many fine nuances. In order to communicate these tones, researchers have come up with a number of different terms such as pragmatic phrases, lexical

phrases, idioms, formulae, lexical collocations, grammatical collocations, rigid collocations, flexible collocations, free combinations, free collocations, and restricted collocations (Granger 147, Jackoson & Amvela 114, Wei 4, Nesselhauf 225, Chang 285). In an attempt to bring some order to this glut of terms, let us begin simply (Cowie provides a more thorough and in-depth discussion of this topic in his essay, "Phraseological Dictionaries: Some East-West Comparisons").

To begin the discussion the reader needs to keep the most general definition in mind: "frequently occurring combinations." As the discussion progresses the categorization of different collocational phenomena will develop into a more defined classification system. Collocations can be divided into three basic categories: free combinations, restricted collocations, and idioms (Hogwarth 164). Free combinations are on one end of the spectrum. Words and phrases that fall into this category combine with other words on the basis of semanticity. For example, literally, *read* is fairly free to combine with many other words as long as they fall into the semantic field of "containing text." One can *read a book, newspaper, magazine, pamphlet, note, brochure, label, advertisement, window* (if there is text on it), *shirt, etc.* If one used a more metaphorical sense of the word *read* the list of words would continue with entries such as *stars, minds, palms, future, etc.* There are many words that would make logical sense if they occurred next to the word *read*. The word *read* is free to combine with numerous other lexical items.

Restricted collocations seem to have a somewhat more arbitrary restriction (Nesselhauf 225). The word *clench* occurs with very few words. One can *clench a fist* or *clench their teeth*, but it would sound odd if someone **clenched their eyes* or **clenched their feet*. The reason that *clench* collocates with some words and not others is unknown and therefore, arbitrary. The same

relationship can be seen with a word like *commit*. One can *commit a crime, commit perjury,* or *commit suicide*, but if someone were to **commit a good deed,* **commit truth,* or **commit revitalization*, something would be deemed incorrect and odd. In the first example, the incorrect forms contain nouns that seem to fall into the same semantic categories as the acceptable forms, but for some reason, *clench* is restricted to few parts of the body. In the case of *commit*, there seems to be a very specific semantic field that it can collocate with, and the contrasting semantic field seems much less acceptable. The preference of words collocating with a limited set of semantic fields is what Xiao & Mcenery have called 'semantic preference' (107). It seems that some words prefer other certain words that fall into the same semantic category. This restrictive relationship is what is known as restrictive collocations.

Idioms can be defined as a group of words (lexical phrase) whose overall meaning cannot be derived from the individual meanings of its constituents. *Kick the bucket* can mean, "to die." There is no way to know that this is the meaning of *kick the bucket* by looking at the meanings of the individual words (unless of course the context provided indicates a literal translation such as: *The boy was so upset with his parents that he kicked the bucket that was sitting on the floor next to the door.*). Although these three clear categories have been provided, it is important to remember there has not been agreement about the criteria distinguishing collocations from free combinations on-the-one-hand and idioms on the other (Bahns & Eldaw 102).

Perhaps better a way to handle this may be to think of free combinations, restricted collocations, and idioms on a continuum, and not just one continuum, but two. The opaqueness of the set phrases and the constriction of the set phrase can be taken into consideration. In dealing with opaqueness, the different types of set phrases fall onto the continuum in much the

same way as they do on the tendency of occurrence, with free combinations on one end and idioms on the other. For example, *read the newspaper, clenched fists, seeing is believing,* and *kick the bucket* move from being semantically transparent to semantically opaque. If one were to focus solely on idioms, this distinction can be seen at a finer level. For example, *seeing is believing* is fairly transparent. Its meaning can easily be derived from its constituents' meanings. An idiom like *throw in the towel* becomes more semantically opaque. One would not know that it means "to surrender" unless she had some knowledge about the sport of boxing. On the far end of the idiom continuum, there are idioms like *kick the bucket* whose origin is unknown, or under debate, and whose meaning cannot be derived from the meanings of its individual constituents.

The other continuum to consider is that of restriction. The restriction of collocations can be defined as the ability of a word to be collocated with other words, whether it is in a syntagmatic or paradigmatic relationship (Moon 95). This type of relationship is similar to what Singleton calls a 'fixed expression' (47). The dispersion of the different types of collocations is not as clear as the previous continuum. When considering the opaqueness of a set phrase, it is a fairly linear progression from free combinations to idioms. The chance of an extremely semantically opaque free combination is very unlikely. The reason that a lexical item becomes opaque is because its meaning cannot be derived from its constituents, and by definition, if this type of opaqueness occurs, one has an instance of an idiom.

The restricted collocation *clenched fist* can be considered highly restricted. In the previous two continuums, a restricted collocation would fall in the middle of the spectrum. In terms of sytnagmatic relations *Clench* is most commonly seen in a V-(DET)-N relationship (*He clenched his fists in anger. He sat there watching with clenched fists.*). When considering the

paradigmatic relationships, there are very few nouns that can collocate with *clench*, and they seem to be confined to the semantic field of *body parts*. To take it one step further, within the semantic field of *body parts*, *clenching* is most commonly done with fists, teeth, jaws, muscles, hands and guts (Corpus of Contemporary American English). So as one can see *clench* as highly restricted. Now let us look at the idiom *drag your feet*. Here you have the same form of V-DET-N (*drag your feet, is dragging her heels, dragging his ass)*. Just like *clench* this idiom seems to have a fairly fixed sytangmatic relationship, and the nouns seem to be highly restricted to and within a semantic field. It would be hard to say which lexical phase would follow which lexical phrase on a continuum. A free combination such as *read + N* would be consider to have a loose or low restriction. The number of acceptable collocates for *read* are many. Granger, who did a collocational study of amplifiers, touches on this topic with his discussion of restricted and open collocability. His use of these terms seems to be more focused on the paradigmatic relations, but nonetheless, it is the same concept (147).

Collocations can also be categorized as 'lexical' and 'grammatical' (Wei 4; Jackson & Amvela 114). When discussing a lexical collocation, one is considering the collocability of content words, the combination of verbs, nouns, adjectives, and adverbs. Certain verbs only collocate with certain nouns that fall into semantic fields, and quite often, the choices are limited within that semantic field (see the discussion of *clench* above). The same can be said with adjectives. It would seem very odd to describe a tree as *passionate*. Usually, *passionate* is an adjective reserved for animate objects that are capable of feeling and emotion.

Grammatical collocations, on the other hand, consist of a content word and a preposition. For example, *refer to* is a highly restricted grammatical collocation. It is hard to think of a

situation where one would *refer in, on, through, about, over,* or *from*. The same can said with *in town*. One can say *Tom is in town*, but it would seem odd to hear that *Tom is on, over,* or *under town*. The *town* example does have a looser restriction than *refer*. Tom could be *near* or *about town*.

Chang describes collocations as being *rigid* and *flexible.* A rigid collocation is one in which its constituents appear adjacent to each other, *refer to* or *freshly baked bread*. If either of these were encountered in context, they would most likely exist as they do here. However, the collocation *a clean slate* may have several different forms in context. One could *wipe the slate clean* or say that *the slate is clean*. In the case of *clean slate*, it is a flexible collocation. The concept of flexible and rigid collocations are related to a word's domain.

The discussion of collocations is not an easy one, and the terms used to describe them are used interchangeably with varying definitions. The information here has been an attempt to simplify some of the basic classifications and continuums for collocations. A thorough look at collocations is provided by Nation. He has identified ten different ways to classify and talk about collocations (329-332). Nation discusses them in terms of how frequent they co-occur, their adjacency (how many words separate the constituents), their grammatical structure, connection, uniqueness, and fossilization, their specialization, their lexical fossilization, semantic opaqueness, and the uniqueness of their meanings. In contrast, Bahns and Eldaw suggest the following simple set of characteristics. Collocations are fairly transparent, they occur frequently, and come to mind easily(102).

The discussion of collocations is as complex, if not more complex, than the discussion on how to define *word*. If we define *collocation* as a group of words that commonly occur together,

then other terms such as *idioms* would be considered collocations. Although idioms are a group

of words that commonly occur next to each other, they have the additional qualifier of their

meaning not being easily derived from the individual constituents. This aspect is not seen with

collocations and therefore, should not be included in the definition of *collocation.* Furthermore,

for the purpose of this study which focuses on ALLs not understanding the full semantic range of

a given word, which precipitates an unnatural word combination, the definition of *collocation*

focuses on the concepts of free-combination and degrees of restrictedness. The definition of

collocation for this study is a group of words that commonly occur because of the grammatical

and/or semantic aspects. Collocations can be very loosely connected such as we saw with *read*

or they can be highly restricted as we saw with *clench*.

1.5 Chunking in Language Acquisition

So far in the discussion of collocations, the terms, characteristics, and criteria have been

explained. So, they exist in language, but why do they exist and what are their importance? The

acquisition of language itself seems to be largely based on the acquisition of language chunks.

Learning language involves the acquisition of memorized sequences, starting at the smallest

units of phonemes to the broader area of discourse (Ellis 93). Ellis gives several reasons for

believing that language is learned and stored in chunks. First of all, native speakers chunk a

language at all levels. A native speaker of a language instinctively knows if a word is from her

language or another. She tends to recognize phonological sequences that do not belong to her

language. A native speaker of English, whose consonant clusters rarely go above three, have a

hard time articulating words from languages with clusters containing four to five consonants.

Many Asian speakers have a hard time pronouncing English consonant clusters because their

languages follow a fairly standard CV structure. Native speakers have an ability to recognize a native sounding phrase. Although "*I wish to be wedded to you,*" is grammatically acceptable, a NS of English would find it odd and would rather say, "*I want to marry you*" (Ellis 97). As one can see, native speakers seem to have a natural ability to recognize chunks of language from phoneme blends to words.

Another example of learners chunking language is given by Ellis (93-94). Ellis looks at the progression of responses in word association studies. The results of his quoted studies are mirrored images of each other for both first language acquisition and second language acquisition. In L1 acquisition, young children, around the age of 6, responded with answers that were phonologically or orthographically similar. Older children began to respond with syntactical responses, providing responses such as *"on-->top"* and "*smooth-->skin".* In the final stage, the responses were related to each other by semantics. The speaker provided responses which fell into the same semantic field as the provided word (Entwisle qtd. in Ellis). As stated before, these results where the same with the L2 learners (Ellis 94).

When one analyzes the results presented by Ellis, a "chunking" pattern can be seen. In the early stages of language acquisition, the learner is grouping the words by what is most tangible and concrete. The learner groups the words by how they sound or how they look. At the next level, the learner provides responses that are bit more abstract and complex. These responses require the learner to know phrasal chunks and syntactical relations between the words. These types of responses are bit more abstract and complex because the learner is now pairing words that have visual or auditorial relations. The knowledge that the learner seems to be relying on is his experiences with the natural use of the language . At the final level, the learner

provides the most complex responses of the three age levels. Responses of the third level require the learner to group words by both denotational and connotational meanings of words. This type of processing moves away from the more linear form of thinking that is present in the syntactical responses. Grouping the words by semantic relations requires the learner to chunk the words by a complex web of relations. In this portion of Ellis's discussion we see language being chunked orthographically, phonologically, syntactically, and semantically.

Another reason for believing that language exists in chunks can be related to research with short term memory (STM). It was found that it was easier to remember a sequence of one syllable words rather than a sequence of five syllable words. This is thought to happen because of limited space of the STM. The more syllables (chunks) to be remembered, the greater the load was on the STM. Furthermore, the processing of these chunks is believed to take place in real time (Gathercole &Baddeley 9). To take this one step further, if the STM runs the risk of being overburdened with chunks of language, and if the brain stored language in small isolated units such as phonemes, or even syllables, the task of creating language would be too great. Therefore, the advantage, and necessity, of storing language in chunks larger than the syllable or word is evident. In order to speak fluently and communicate, whole phrases of words would need to be stored in mental lexicon, not just single words (Ellis 107). Not only would chunking help the speaker, but also the listener. The more the units of language come as packaged wholes, the greater the possibility of attentional focus (Ellis & Sinclair 245; Ellis 111; Hyland 43). Ellis and Sinclair go on to say the STM thinks of lexical phrases as large words. They do acknowledge that that the complexity of phrases does create burden on the STM, but just as the STM aids in the retention in the longterm memory of vocabulary, it does the same for phrases (245). Nation

echoes these thoughts by saying, "language is possibly chunked as a way of saving on processing time" (319).

It seems evident that chunking occurs naturally in language acquisition. It happens with the smallest components of language up to the sentence level. In order for our minds to efficiently create smooth and seamless language, it needs to be able to store larger chunks of language. As stated before, the idea of storing language in prefabricated chunks does not inhibit a language's creative ability. The constituents of a chunk are still able to be moved, substituted, and manipulated. These movements, manipulations, and substitutions are evident in collocations. Because of this flexibility, native speakers rely on grammatical and semantical knowledge to determine the what words are acceptable to collocate.

There has also been research done using large corpora and concordances. Melcuk found that a 'phraseme' or 'set phrase' is the numerically predominant lexical unit, and it outnumbers the words roughly ten to one (24). Howarth, whose corpora consists of nearly 250,000 words, suggests the possibility that there is a "conventional core of collocational conventionality" in NS writing. About 35% of the text analyzed consisted of restrictive collocations. This number in itself is significant, but then also consider Butler who found that in comparing written and spoken corpora that collocations in spoken language are more frequent than in written language (qtd in Ellis 97).

Shin and Nation conducted research on collocations in spoken language. They used the British National Corpus, which contains approximately 10 million words. They were looking for collocations containing high frequency pivot words (noun, adjective, verb, or adverb). With the help of a computer program that was able to isolate the pivot words, they reported four major

findings. First, there was a very large number of grammatical well-formed high frequency collocations. A total 4,698 collocations were found whose pivot word was found in the top 1000 frequently used words. Second, the more frequent the pivot word, the greater the number of collocates. Third, a small number of pivot words account for a very large portion of the tokens of collocations. "When the frequencies of the all the collocations were added together, the total number of occurrences of the collocations of the first 100 pivot words is 387,634 which covers about 53 percent of the total number of tokens of the collocations found (736,144) (Shin & Nation 343-344). Finally, they found that the shorter the collocation, the higher the frequency.

There seems to be sufficient evidence to support the hypothesis that language is used and stored in chunks. Some may say that the idea of a prefabricated language runs contrary to the idea of language being free and creative, but this does not need to be so. The idea that language is prefabricated is accepted at the morpheme and word level. Therefore, one only needs to expand his acceptance of what a word is and the ability of being free and creative maintains its integrity. If an individual has an active vocabulary of 50,000 and half of that consists of prefabricated chunks, the theoretical ability to create unique utterances teeters on the edge of uncountable.

1.6 Collocation Studies in SLA

Recent research has shown that collocations are a continual struggle for language learners. Second language learners are often not exposed to the full range of language and its registers, they are not often fully exposed to a language's colloquial use, and they struggle with opaqueness of idioms and the heavy connotative meanings of some words. Nesselhuaf conducted an exploratory study that looked at the use of lexical collocations, more specifically verb-noun

collocations, and some possible contributing factors. His research was based on a German subcorpus ICLE (The International Corpus of Learner English). The study looked at thirty-two essays from German students in their third to fourth year. Once the essays were collected, all of the verb-noun combinations were extracted, classified by their degree of restriction, and then, judged on acceptability by native speakers.

Nesselhuaf found that nearly one-fourth of the identified V-N collocations were rated as wrong by native speakers. There was also weighted distribution of the errors to the free combination side. That is, the looser the restriction of collocability the more likely a learner was to make a mistake in its use. " It therefore seems that whereas learners are mostly aware of the restriction in combinations where the verb takes only a few nouns, they are less aware of the restrictions in combinations where the verb takes a wider range of nouns" (233).

Nesselhauf also looked at the possibility of L1 influence on the collocation mistakes. Although there is no certainty that L1 made an influence, Nesselhuaf took into consideration the similarities between the L1 form and the produced form. "If *make homework* was produced, the fact that German has *Hausaufgaben machen* and that *machen* is related to *make* in both meaning and form led to the assumption of L1 influence" (234). He concluded that there seemed to be L1 influence on all types of collocation mistakes (235).

Bahns and Eldaw also conducted collocation research. Their focus was similar to Nesselhuaf's in some ways and different in others. Both Nesselhauf and Bahns and Eldaw focused on the verb-noun combination, and their German subjects were considered advanced learners, but this is where the similarities end. Bahns and Eldaw used a cloze procedure and translation task to test the learners' ability to produce appropriate collocations. Also, Bahns and

Eldaw's research focused on whether or not students will simply acquire collocations without explicit instruction, and if their collocation knowledge expands with their general vocabulary knowledge. The study also looked at whether or not collocations were necessary for effective communication, or would the ability to paraphrase around an unknown word be sufficient. Finally, they aimed to discover if there was a systematic way to choose appropriate collocations to teach (103).

Fifty-eight German students participated in the study. Thirty-four performed the translation task and twenty-four performed the cloze task. The subjects had to actively produce fifteen V-N English combinations. In the translation task, the subjects were given the German equivalent of the V-N combination.

Bahns and Eldaw concluded the following. First, a learner's knowledge of collocations does not expand parallel with general vocabulary knowledge. Bahns and Edlaw state, "It can be concluded from this study that learners are more than twice as likely to select an unacceptable collocation as they are to select an unacceptable general lexical word, and that EFL learners' knowledge of general vocabulary far outstrips their knowledge of collocations" (108). Second, collocation knowledge is necessary for full communicative mastery of English. They came to this conclusion after noticing that the subjects failed to successfully paraphrase the target item. Therefore, if one were to argue that successful communication can be reached with paraphrasing unknown items, the data in this study would refute that claim.

Granger's research is more focused than the previous two studies. Granger focused on the use of amplifiers (*perfectly*, *closely*, *deeply*, etc) and their collocates, and the use of these collocations in both native and nonnative speakers' essays. Granger hypothesized that native

speakers would use a greater number and variety of these amplifier collocations. Granger used a French subcoprus of the ICLE corpus for the nonnative portion and a combination of combination of the Louvain essay corpus, the International Corpus of English, and the Lancaster-Oslo-Bergen (LOB) corpus for the native portion of the study.

Granger's data showed that nonnative speakers significantly underused the amplifier collocations in comparison with their native speaking counterparts. After comparing the number of types of tokens in both copora, there was significant underuse in both the numbers of types and the number of tokens used (147). Furthermore, Granger found compelling evidence that showed the learners' L1 strongly influenced their choices of amplifiers and the words they collocated with in the essays. The learners heavily relied on the amplifiers *completely, highly,* and *totally*, and their use of these amplifiers, for the most part, were found acceptable by native speakers. However, Granger points out that all three of the amplifiers have a very close semantic and collocational distribution in French, which is perhaps why the learners felt so comfortable using them (148).

Howarth, like the others, focuses on verb-noun combinations. His study aims to investigate the occurrence of collocations in native and nonnative academic writing. He believes that 37-46 percent of texts can be classified as restricted collocations or idioms. The native speaker data comes from the LOB corpus and a series of text donated by Leeds University. The nonnative data comes from ten essays written at the end of the term of a one-year masters course in Linguistics and English language teaching. The two types of data were analyzed in two different ways. The native data was analyzed in the same way that Nesselhuaf conducted his research. The V-N combinations were extracted from the text and then classified as a free

combination, restricted collocation, or idiom. The nonnative text analysis was done on paper.

They too were extracted and classified.

Howarth found that 33% of the total analyzed texts consisted of restricted collocations,

62% were free combinations, and 5% were idioms. Nonnative texts consisted of 24% restricted

collocations, 69% of free combinations, 1% idioms, and 6% miscollocations. The percentages

found by Howarth are within the figures provided by Cowie (qtd. in Howarth 171). In regards to

the nonnative data, Howarth states that native speakers employ about 50% more restricted

collocations than nonnative speakers (177).

Looking at the research in this section, it appears that native speakers naturally use

collocations in their language. Some collocations are highly restricted such as idioms or

transitional phrases, while other collocations are less restricted such as those found in Howarth's

research. On the other hand, language learners continue to struggle with with both of these items.

Considering the research on language chunking and short term memory done by Gathercole and

Baddeley, perhaps the language learners are not storing these collocations as chunks, but rather

as independent constituents. Regardless of the reason, it is evident that native speakers use

collocations to make their messages clear and concise, and second language learners struggle

with them.

1.7 The importance of Vocabulary

For most of linguistic history, the unit of the word has been largely overlooked. Much time has

been spent on analyzing the phonetics and phonology or the syntax and grammar of the language

(Llach, *Relationship Lexical Error and Composition* 1). Interestingly, vocabulary has been

identified as one of the most important factors of language by many researchers (Nation 145;

Shahheidaripour 3). Saville-Troike et. al did an extensive qualitative study on the different

factors that made second language learners (SLL) academically successful in the English as a

second language (ESL) classroom and came to the conclusion that vocabulary was the most

important factor in order for ESL students to be academically successful (216). Saville-Troike

ended her study by stating, "we must begin to place more emphasis on vocabulary learning and

less on grammar and pronunciation."

  This statement may seem bold to some, but when one considers the function and purpose

of the word, it is quite logical. Singleton presents several perspectives including the "London

School" perspective, the Valency Grammar perspective, and the Lexical-Functional, which all

see the word as the core of a language (20-23). Essentially, the different perspectives assert the

notion that the individual starts with a concept and then searches his mental lexicon for the

appropriate lexical item. Once the appropriate item has been chosen, the grammar for the

utterance is created. For those strong Chomskians, Singleton points out that this line of reason is

not contrary to Chomsky's perspective and that as Chomsky's theories have evolved, it seems to

reinforce the importance of the word in language. O'Rourke believes that vocabulary is

important because we use words in order to communicate and understand ideas (94). The Sapir-

Wharf hypothesis suggests that we perceive the world differently by the words in our language,

and although there is quite a bit of grey area in this hypothesis, it does ring some truth. For

example, if one is learning a language and he encounters a concept such metaphysics, he will be

unable to understand or communicate about this idea if he does not have the proper vocabulary.

"Expansion of vocabulary generally accompanies expansion of thought" (O'Rourke 28).

From a less theoretical point of view, many researchers have commented on the importance of vocabulary. Nagy and Anderson suggested that teaching vocabulary can improve reading comprehension, writing skills, speaking vocabularies, and scores on standardized tests (27). Others agree with Nagy and Anderson about reading comprehension (O'Rourke 25; Hirish and Nation qtd. in Nation 147). Nation is able to put statistics to this point of view, "a learner needs know 95% of the words in a text in order to have reasonable comprehension" (114). Although the focus of previous information has been on the influence of vocabulary on reading comprehension, it is easy to see possible implications for the other skills. For example, a distinction can be made between the receptive and productive vocabulary. Hatch and Brown define these two terms as follows: *receptive vocabulary*—"words that students recognize and understand in context, but which cannot be spontaneously produced," and *productive vocabulary* —"words which students can pronounce correctly and use constructively in speaking and writing" (370). In light of this information, the relevance of vocabulary in the other skills can be seen. It has been shown that vocabulary plays a significant role in the learner's reading comprehension, which is a receptive skill; therefore, it is not too far of a stretch to extend this logic to the other receptive skill of listening. It is reasonable to say that an individual would need to know 95% of the words heard in order to derive accurate meaning from the interlocutor.

If we follow Nation's suggestion that receptive learning is easier than productive learning, then the receptive vocabulary will be larger than the productive vocabulary (32). Laufer supports the idea of the receptive vocabulary being larger than the productive; furthermore, as to be expected, she found the passive (receptive) vocabulary developed at a faster rate than the active (productive) vocabulary (12-13). Thus an obvious importance of vocabulary can be

attached to all four skills. Both reading and listening will require a vocabulary that covers 95% of the input to extract meaning, and an even better productive vocabulary will be needed to create meaning through writing and speaking. To take this one step further, Nation asserts a need for a larger writing vocabulary than speaking vocabulary because we tend to write about a wider variety of topics than we speak about (125).

The learner's vocabulary in both their L1 and L2 is an intricate part of his language success. In academia, the learner needs to have a strong lexicon in order to be successful at listening and reading comprehension, and speaking and writing production. In colloquial use, the learner must be able to access the correct lexical item in order to communicate the correct idea. Although it is true that an individual needs to be intelligible, and he also needs to be able to use correct grammar, it is the word that carries the larger portion of the meaning of the utterance.

2. Lexical Errors

With an importance of vocabulary established, one needs to decide what to do with this knowledge. How can educators improve their instruction with the importance of vocabulary in mind? One approach is to look at the types of lexical errors that learners are making. Are they misusing certain parts of speech, are they making errors due to the semantics, are they making errors because of L1 influence? Once the types of errors are identified, methods and activities can be created that address the identified problems. Several scholars have already done research on lexical errors in L2 production.

To preface the discussion on errors, in Nesselhauf's study, he points out that "despite the attempt to be precise as possible, the classification presented has a number of limitations...word combinations differ along a scale which makes their exact delimitation impossible" (227). So,

although I will attempt to set precedents for identifying and classifying lexical errors for this study, there will be some vague and unclear limits between error occurrences and how they are classified. To help make these boundaries clearer, we can use Lennon's definition of an error, "a linguistic form or combinations of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speaker's native counterparts" (182). Further on in his study, he discusses errors in terms of "error extent" and "error domain" (191). Error extent considers the rank of the linguistic unit. It looks at whether the error occurs anywhere between the morphemic, lexical, or syntactic level. It also includes whether it is a derivational or inflectional miscue, or whether a whole sentence needs to be deleted, replaced, or reordered in order for it to be intelligible and comprehensible. Error domain is a linguistic unit which must be taken as context for the error to become apparent. A word in isolation may not seem erroneous. *Mate*. It may also seem acceptable in a phrase *a mate*. However, in the context of the complete sentence one might find the word *mate* a bit abnormal. *He found a mate to dance the waltz with him*. In which case, perhaps a more appropriate word would be *partner*. In the following reviews of lexical error research, each researcher has slightly different error domain and error extent. The hope is that by analyzing the different approaches the current study will find some overlapping area that will lend itself useful for identifying and classifying the errors in the current data.

2.1 Types of Lexical Errors and How They Affect Language

Zimmerman has attempted to extensively describe the different types of errors that come from German university and high school English language learners (ELL). Zimmerman expresses his frustration in the oversimplified classifications of previous works (Classification and Distribution

31). He attempts to improve the traditional classifications of *lexical, collocational,* and *idiomatic*

errors (see table 1.1).

Table 1
Error Examples from Zimmerman's Study

Confusion of Supernyms and hyponyms in either direction	Hyp→Sup *smell* instead of *scent* Sup→Hyp *claims* instead of *demands*
Erroneous Supernyms that are also Stylistically Inappropriate	*eradicate* instead of *Getting rid of*
Co-hyponyms have been Confounded	*eradicate* instead of *exterminate* (both are "terms of forceful removal…, but the one does not exclude the other." (p33))
Field Errors	*memorandum* instead of *memorial* ("*memorandum*, a word from the semantic field of 'written public communication' [and] *memorial* which is from the field of 'public buildings'" (p33))
Feature Error (synonymy)	*slipped* instead of *slide*
Word Formation Error	*yellowly*
Collocation Error	*red-rimmed eyes* instead of *red-edged eyes*
Noninterpretable forms	*memorandum* instead of *momentos* *clops* instead of *lumps*
Idiomatic Errors	*Jack-of-all-trades* instead of *Jack-in-the-pot*
Redundancy	*At a rate speed of…*
Omission	*I had bitten through (the edges of) my tongue.*
Paraphrase	*Outer parts* instead of *edges*
Stylistic Errors	*Pubs* instead of *taverns*
Connotative Errors	*Jumped* instead of *jerked*

Zimmermann's results showed that 20% of the errors could be classified as sense relation

errors, 33% as field errors, 10% as feature errors, and 20% were seen in the wrong collocation

(37). Zimmermann also notes that paraphrases play a large role in his data and that they occur

more in spoken utterances as a result of less planning time on behalf of the speaker, but he

suggests that paraphrasing be encouraged as a navigational strategy.

In Zimmermann's later work, he posits that learners' lexical errors can be largely categorized into form-orientated and content-orientated errors (*Form and Content Orientated* 55). Zimmerman's methodology is focused on translations activities to collect his data. Using data from advanced German ELLs he gives the following examples for form-orientated and content-orientated errors (see table 2).

Table 2
Examples of Form-Oriented and Content Orientated Errors from
Zimmerman Study

Intended word: *sledge track*

Form-Orientated	Content-Orientated
slide track	toboggan track
slade track	ski track
sledge-spur	skate track

Zimmerman then suggests different reasons for error occurrence. He suggests that most form-orientated errors are the result of the learner attempting to make an approximation because he/she does not know the word. The learner takes into consideration the graphic and phonetic characterizations of the target lexical item in their mother tongue or the target language. As seen above the target word was *sledge*. All of the given words have similarities in the graphemes or the phonemes used. The second classification Zimmermann uses is content-orientated errors. He defines them as, "forms which are not form-orientated (55)." Zimmmerann's definition of *content-orientated* is fairly simple and imprecise. It seems that his *content-orientated* classification is closely related to, if not the same as, misusing words in the same semantic field. He suggest that the learner chooses a word from the same semantic field in order in attempt to make an approximation of the target word (61). As with most classifications, Zimmerman acknowledges that there are areas of overlap in his taxonomy (64). Although Zimmerann's

methodology focuses on direct translation activities and not the analysis of naturally produced

language, his classification presents an alternate perspective and is worth consideration (see table

1.3). There is not a table for the content-orientated errors because of Zimmerann's definition

(anything that is not form-orientated). The word given before the arrow is the word used by the

learner and the second word is the target word in the second language (L2) (see table 3).

Table 3
From-Oriented Error Examples from Zimmerman's Study

Form-Orientated	
Phonic-Graphic L1 Orientation (false-friend)	*ringkampf→ring-fight*
Phonic-Graphic L2 Orientation	*Ansporn (incentive)→incitement*
Indirect Form Orientation	*scheinend→shining*
Phonic-graphic Orientation Ambiguous between L1 and L2	*schrammte (scraped)→scrabbed*
Morphological L2 Orientation (overgeneralization in word formation	*Side-character→minor character* Created because of such words as *side-effect*
Phrase Level Form-Orientation	*Handsdampf-in-allen-Gassen (Jack-in-all-streets)→Jack-of-all-trades*

Meara and English also studied lexical errors, but they focused on the lexical errors and

the use of learner's dictionaries. They used a corpus of 1364 errors that came from a collection of

First Certificate examination papers (2). They identified 14 different language groups in the

corpus. Meara and English attempted to determine the usefulness of a Longman's learner

dictionary as a tool to correct errors in writings. They first separated the errors into different

fields: look-up words, a short context, a source language code, an error type code, and a

dictionary code. Then the fields of *error type code* and *dictionary code* were further divided into

subfields. The error type code is the one of interest for this study. Meara and English used the

following coding system:

0   totally wrong word

1   phonologically related word

2   wrong word right semantic area (errors of style and register)

3   formal derivational errors (words basically correct, but wrongly formed)

4   usage (the word is correct, but its context is incorrect)

5   spelling errors

It was found that error types (2) and (4) made up more than 50% of all the errors in the corpus. Both of these errors could be reclassified as semantically based errors. There were some variations in errors types between languages. Most noticeable was that Chinese and Indonesian. They were the only two languages to have more than 10% of their errors fall into error type (1) with 24% and 52% respectively. Meara and English do recognize that generalizations cannot be made from their study alone.

Salem takes a slightly different approach to lexical error analysis. Salem focuses on the relationship between the lexis and grammar and how different combinations of the two affect comprehensibility. This approach comes from the view that "the binary distinction between grammar and lexis has…been challenged by the lexical approach, which proposes that the basis of language is lexis" (212). He also uses the idea of "word sensitivity" in errors, that severity of some errors is affected by lexis more than other types of errors. The study was conducted by taking 20 errors from Israeli high school EFL compositions. The 20 errors were judged by 22 EFL teachers working at the high school and 11 native speakers (NS) of English working as TESOL/linguistic university teachers. The following is a classification of the errors;

word selection error—*light* instead of *easy*

inaccuracies in tense morphology and usage—*She had gone just now*

collocation errors

negative transfer errors

Salem further divides the errors into three broad groups of lexis, word-dependent, and pure grammar. He suggests that each field has a different level of grammaticality and word sensitivity. *Word sensitivity* is essentially a words ability to co-occur with another word and the relationships between the two words that allow them to occur together (213). For example, one can "ring a bell", but cannot "ring a box" because "ring" is restricted to certain nouns that have the ability to resonate in a way that creates a higher pitched ringing sound. The requirements that need to be met for "ring" and "box" to occur together have been violated. This is the idea of word sensitivity.

 In the field of lexis, he has collocations on the lexis end of the spectrum, word form on the grammatical end, and word choice somewhere in the middle. In the same field, collocations have low word sensitivity, word choice has mild word sensitivity, and word form has high word sensitivity. Salem breaks down the other two fields in similar ways. He suggests that if teachers are aware "of the distinction between highly and mildly word-sensitive errors, [it] might become instrumental in responding to learners' requests for error clarification" (218).

One final consideration is Llach's study, which looks at lexical errors as writing quality predictors. The study involved the compositions of 71 fourth grade Spanish EFL learners. Llach developed four main categories of lexical errors (see table 4).

Table 4
Lexical Error Categories form Llach's Study

Misspelling	"violations of the orthographic conventions of English" i.e. *biutiful*
Borrowings (complete language shift)	"a learner inserts any L1 word into the L2 syntax" i.e. *My grandmother is **coja**.*
Coinage	"the adaptation of an L1 word to the L2 orthography or morphology" i.e. *My rabbit is small, very **divert*** (Sp. Divertido; Eng funny)
Calque (literal translation)	"when a learner literally translates the word from L1" i.e. *table study→mesa de estudio*

Llach found that 78% of the errors were spelling errors, although they had little impact on the overall quality. Borrowing constituted 10% of the errors and had a greater impact on the composition quality. Finally, it was found that coinages were 9% and calques were 6% of the errors and both, similar to the misspellings, were not very effective at predicting writing quality (*Quality Predictors* 9). Llach surmises that lexical errors are not effective criteria for predicting writing quality when communicative effectiveness is the main criteria for the compositions (*Quality Predictors* 15). Llach does suggest that further research is done to see if the lexical hierarchy holds true for higher leveled learners. In addition, although Llach's study used fourth grade students as study subjects, his classification system in relevant to the current study.

To conclude, several different approaches to classifying lexical errors and their affects on language have been presented. Others in the field have come to many of the same conclusions as the presented research with some additional thoughts. Nation attributes many lexical errors to the fact that L1 words may have competing associations with the target word in L2 (29). He also suggests that using dictionaries can result in the misinterpretation of a word's meaning (83).

Hatch and Brown believe that many errors occur because learners confuse them with words that have similar forms in their L1 or L2 (378).

Finally, in their research, Xiao and McEnery focused on the possibility that both Chinese and English have both collectional or semantic preferences the with following semantic groups: consequence, cause, and price/cost. They looked at the semantic preferences for these group in both English and Chinese. They found that there is both overlap and divergence in the semantic preferences with these words. Xiao and McEnery posit that semantic lexical errors may be the result of confusion with semantic preferences with words. The learners transfer their knowledge of L1 semantic preferences to the production of their L2, resulting in nonnative-like utterances. These findings have influenced the current study to look at miscollocations as possible reason for learners' semantic lexical errors.

3. Vocabulary Instruction

Traditionally, vocabulary instruction has received little attention in both L1 and L2 instruction. In the L1 classroom, there is between .5 and .8 percent of instruction spent on vocabulary development (Baumann & Kameenui 611). Vocabulary instruction usually receives more attention in L2 instruction, but traditionally, it has focused on memorizing a list of words and treating them as isolated units (Robinson 276). Recent vocabulary acquisition and instruction research has focused on a deeper and broader approach to L2 vocabulary instruction. Linguists, curriculum developers, and instructors are aware that there are many aspects to knowing a word, not just their written/spoken form and the denotata. O'Rourke states,

> [Vocabulary instruction] involves problem solving and the use of [the learners'] mental
>
> processes. It includes encouraging (and providing the opportunity for) the student to

observe, compare, classify, interpret, hypothesize, criticize, and create. In short, the

student is asked to learn about words and learn words in a variety of ways aimed at

forming concepts. (94)

In the view of O' Rourke and others, vocabulary instruction has moved beyond the simplistic

instructional styles that incorporate activities such as memorizing vocabulary (Nation 23; Hatch

& Brown 373; and Shahheidaripour 4).

Vocabulary instruction can be broken into two parts, how to teach and what to teach.

There are several considerations when one is considering how to teach vocabulary. First, let us

consider the receptive and productive vocabulary. The definitions and differences between these

two terms have already been discussed. In general, a person's receptive vocabulary is larger than

his productive vocabulary. This is probably because receptive vocabulary and receptive learning

are easier than their counterparts (Nation 28, 32).

Another consideration is the form of vocabulary instruction. Should it be implicit or

explicit and should it be inductive (incidental or indirect) or deductive (intentional or direct).

Explicit learning happens when the learner is aware of what is to be learned. Implicit learning is

learning without being aware of what is to be learned. Inductive learning is moving from the

general to the specific, and deductive learning is moving from the specific to the general

(DeKeyser 314; Hatch & Brown 368). Dekeyser explains a framework for these two dimensions.

Instruction can be explicit and deductive, explicit and inductive, implicit and deductive, or

implicit and inductive (314). Giving learners a list of vocabulary words with their definitions

would be considered an explicit and deductive method. The teacher is making the students aware

what vocabulary they are to learn, and the teacher is providing them with a specific definition.

When a teacher provides the learners with a list of vocabulary and has the learners define the words using the context, then the teacher is using an explicit inductive approach. Once again the learners are aware of the vocabulary to focus on, but in this case, the learners have to come up with their own definitions. The other two, implicit deductive and implicit inductive are harder to explain. An implicit inductive example is the natural acquisition of vocabulary. Learners are not aware of the words they are learning, and they are creating meaning from their own experiences with the word. In SLA, extensive reading is a type activity that would promote this type of vocabulary acquisition. An example of an implicit and deductive approach does not exist in vocabulary instruction. When Dekeyser describes it, he offers the setting of parameters (referring to Chomsky's theories) as an example of implicit deductive learning (315). To the knowledge of this researcher, there is no part of vocabulary that coincides with concept of parameters in vocabulary acquisition.

This, like other topics, is not a dichotomous split. Opinions on implicitness, explicitness, directness and indirectness exist on many points of a continuum. There are those who support explicit approaches. For example, Nation feels that aspects related to meaning should be taught explicitly and that explicit learning draws attention to the item and helps to create a mapping or linking of knowledge (34). Nation also believes there is a need for the direct instruction of high-frequency words for SLLs. He points out that the average native speaker comes to school with a vocabulary that contains around 5000 word families (156-157). A word family consists of the head word and its inflected and derivational forms (8). This vocabulary base contains the high frequency words of the language; therefore, there is no need to directly teach high frequency words to NSs. On the contrary, if the NNS comes to school without that vocabulary base, Nation

feels that there needs to be direct instruction of those high frequency words (126). Those who are

against explicit deductive instruction will assert that there are too many words to teach and that it

is futile to try and directly teach them. Supporters of indirect learning will site reading as a good

example of vocabulary acquisition as a result of indirect learning (Dupy and Krashen qtd in

Hatch & Brown 369). This stands in contrast to Nation. Indirect supporters might assert that the

learning of the high frequency words would naturally occur through large amounts of level

appropriate reading. The actualization of the issue is that there is no clear answer about what type

of instruction is best all the time, which is why it is suggested that there be a balance among the

four approaches (O Rourke 25; Baumann and Kameenui 609).

3.1 Principles of Vocabulary Instruction

With the different views and opinions of how vocabulary should be presented there are various

types of methods and strategies that coincide with these views. Not every method and strategy is

appropriate for every situation, but they all do have their uses. Before specific methods and

strategies are discussed, some basic principles of vocabulary instruction need to be considered. It

is also important to consider what type of vocabulary should be taught.

It seems that noticing the words and repetition of the words are two reoccurring strands

found in vocabulary instruction literature. Hatch and Brown discuss *noticing* in terms of

encountering words or having a source for new words (373). In direct instruction encountering

new words usually happens with a words list. The teacher may give the students a list of words

to translate and memorize. Or the teacher may give the students a list of words to look up in the

dictionary. On the converse, encountering new words in incidental learning happens through

natural interaction with the language community. The learner listens to the TV or radio and reads

newspapers, magazines, and other realia. Incidental learning can be focused by the teacher. She may have the students read a book and keep track of words that they did not know. Hatch and Brown site several pieces of literature that support the use of incidental learning as the most productive way of learning vocabulary. Hatch and Brown feel it is important to point out that with either direct or incidental learning, an important factor of acquiring a lexical item is whether or not the learner has an interest in the item (373). There may be several reasons that a learner may take an interest in a new lexical item. A few reasons are needing to know the meaning to navigate an oral interaction, needing it to understand a written a passage, or simply having a personal interest in a certain semantic area which naturally draws them to a certain semantic area.

Nation defines noticing words as a learner encountering a word and recognizing it as a useful language item (63). Nation's view does not differ much from Hatch and Brown's. In Nation's discussion, the concepts of incidental and direct learning are not a major focus. Nation simply states that the word needs to be identified by either the learner or the instructor as an important item. This could be interpreted as the dichotomy of incidental and direct instruction that was identified by Hatch and Brown. Nation goes on to give different methods the leaner can use in order to derive meaning or classify a new word once it is encountered. Although Nation's semantics may differ from Hatch and Brown, he agrees that noticing words can happen in many different ways. Nation gives the following possible ways of noticing a word. A learner can recognize a word as new or different while reading and listening, having a teacher highlight a written word on the board, negotiating meaning in conversation, having a teacher explain or define a word, give a synonym, or translating from his first language (64).

Finally, O'Rourke weighs in on the topic by saying that it is important to "get students to notice words and pay attention to their meaning and structure" (86). Although all three sources have a slightly different approach to the concept of noticing a word, they seem to agree on the importance of the learner being exposed to new vocabulary and having some sort of vested interest in knowing the word.

Another common principle of vocabulary instruction is repetition. It is agreed that a learner needs to have multiple exposures to the word. It is thought that a learner needs to encounter a word between six and twelve times in order for it to be retained (Dixon qtd. in Hatch and Brown 375). However, several people believe that it is not as simple as encountering an item in passing. It needs to be encountered in multiple sources and contexts (Hatch and Brown 375). Nation believes in the importance of multiple exposures in different contexts because a word has so many different grammatical and semantical aspects. The more the word is repeated in instruction, the better chance the learner has to retain it (76).

The idea that a leaner needs to have multiple encounters with a word in different contexts has been quantified in the "involvement load hypothesis." It was first studied by Hulstijn and Laufer in 2001 and again in 2008 by Kim. The involvement load hypothesis is a way to measure motivational and cognitive involvement in a vocabulary task. It is comprised of three parts: need search and evaluation. *Need* is based on how important the item is needed to complete a task, whether it is a formal instructional task or a conversation. *Search* is whether a learner is required to search for the definition in a dictionary or some other source such as an interlocutor or context clues. Finally, *evaluation* is the learner making decision about the lexical item such as how it compares with its other meanings or how it interacts and occurs with its surrounding words.

Kim, who replicated the original study by Hulstijn and Laufer with an additional component, found that those vocabulary tasks that had a higher level of involvement were more beneficial than those tasks with a lower level of involvement (311). This finding was in line with Laufer who concluded, "that the chance of a new word being stored in long term memory is determined by the depth at which the word is processed" (qtd in Kim 287). This supports the previous assertion about repetition. Learners need to encounter and interact with words many different times and in many different ways. They need to see the lexical item in different contexts so that they can acquire its different meanings and nuances. Nation gives some tangible examples of the different forms of repetition (80). Repetition can come in the form of notecards, but this does not have to be a repetitive oral phonological repetition of the word. For example, learners can create different sentences with the cards. Repetition can occur in conversations and in reading. It can come in the form comprehensible output, and the repetition can be productive or receptive.

Many researchers agree that vocabulary instruction has to go beyond the traditional memorization of word lists (Kim 311; Wei 1; Shahhaidaripour 5; Nation 96; Hatch and Brown 373). Kim suggest looking at the *involvement load* and Nation advocates, amongst other things, the use of *rich instruction*.

Rich instruction is synonymous with direct instruction. Nation gives pros and cons for using rich instruction. Arguments against rich instruction are that rich instruction can be time consuming, there is too much information to learn about each word, there are too many words to teach directly and there are more effective and efficient ways to teach a large amount of vocabulary, such as incidentally through reading and speaking. However, Nation also points out

that there are instances when rich instruction can be beneficial for second language learners. Language learners need to know high frequency words and since the number of high frequency words are fairly small, direct instruction is feasible (high frequency words include high frequency words specific to certain areas of study, i.e. academics, chemistry, engineering). He also points out that second language learners need large amounts of contact time with the target language in order for them to incidentally learn words. This in itself is not feasible for most language learners, so direct instruction of some vocabulary may be the only way for the learner to get an accurate representation of the word. Finally, Nation points out that second language learners have less time to learn a language. Most second language learners start later in life and are past the critical period, so the acquisition of vocabulary needs to be more focused (95-97). Overall, Nation concedes that rich instruction needs to be used as part of a balanced language program with a limited amount of vocabulary and appropriate time constraints (97).

Nation describes rich instruction (97-98). It can involve having students use concordances to look at the contextual surroundings of a word. By making the focus narrower, students can study the specific collocations of words. For example, if a second langauge learner is taking a Business English course, he would be able to take vocabulary specific to that area and look up its collocations. He would then be able to use that data to make judgements on the types of words that naturally collocate with a specific vocabulary item. Students can also do semantic mapping with a given lexical item. This can easily be connected to the collocational relations between words, looking at the register of the word, if it is related to a specific semantic area, and how it is related to the semantic areas that share certain aspects of the direct semantic area. With

this, several definitions of a word can be looked at in order to identify commonalities between definitions, or to identify specialized contexts of the word.

For example the word *vessel*. *Vessel* could be considered as a word in the formal register. The informal version of the word might be *ship*. It is most often used in the semantic area of "ways to move goods." Metaphorically, it can be used to convey the idea of dissemination or transfer of information. The dictionary definitions would give further information as a *vessel being* a hallow receptacle that is used to carry liquids, a ship or flying aircraft, used medically in terms of blood vessels, or as somebody being a vessel of knowledge or vessel in which knowledge is poured. With this information and the use of a concordance and collocations, a learner could have a strong idea of when to appropriately to use the term *vessel*. For example, knowing that the definition of a vessel includes ships and boats, a learner would know that referring to a bicycle as a vessel may be in appropriate. Knowing that *vessel* is a formal word, it may deter them from using *vessel* when asking for a fishing boat in rural areas of the northern Midwest of America.

Although both of these considerations are helpful tools in considering the creation of activities, they do not provide an easily applied step-by-step procedure for vocabulary instruction. On the contrary, Brown and Payne have come up with five essential steps in vocabulary learning:

> (1) having sources for encountering new words, (2) getting a clear image, either visual or auditory, for the forms of the new words, (3) learning the meaning of the words, (4) making a strong memory connection between the forms and meanings of the words, and (5) using the word. (qtd in Brown and Hatch 373)

As discussed before, encountering new words means the learner is exposed to new lexical items and identifies them as items that are important for comprehension and communicating. The next step, *getting a clear image*, involves the learner learning the form of the word, either written or auditory. Without a clear written or auditory image of the word, the leaner will often confuse the target word with other similar L2 and L1 words (Hatch and Brown 378). In terms of the auditory form, this makes sense. When learning a new word, the learner uses previous vocabulary knowledge to learn the phonological form of the new word (Gathercole and Baddeley 52). Without a clear image of the phonological form, the learner can easily mistake the previous known lexical item for the new target lexical item during the retrieval process. In regards to the written form, children who are learning to read in their native language will often produce inaccurate attempts in instances where a new target word resembles a previously known word. It stands to reason that the same pattern could be found for those learning a second language. It is important for the learner to create strong connections between the words and their forms.

The third essential step is *getting the word meaning*. This involves the learner knowing both the denotational and connotational meaning of the word. Hatch and Brown suggest that the level of "knowing" changes with the level of the learner and the communicative situation (382). The production of an inaccurate word would be more acceptable by a beginner language learner in a casual conversation; on the contrary, it would be less acceptable by an advanced language learner in a formal presentation. In terms of a final outcome, learners should strive to learn the different meanings of a word and its collocations as they progress in their proficiencies.

The fourth essential step identified is *consolidating word form and meaning*. This involves strengthening the mental linkages between the form and the meaning. This can be done

with semantic mapping, building on schemata, connecting old and new phonological forms, and many other strategies (387). Nation agrees with Hatch and Brown, "learners need a strong connection between form and meaning in order to quickly access the word and accurately know it" (48). It connects back to the idea of exploring a word in multiple contexts and perspectives. The fifth essential step is *using the word*. This helps move the new lexical item from the receptive to the productive vocabulary. It also gives the learner the chance to explore and negotiate the meaning and appropriateness of new lexical items.

These five steps provide a useable format for creating vocabulary instructions. It is also a good representation of the basic principles of vocabulary instruction. It incorporates the important ideas of noticing words, getting multiple exposures to them, and understanding the different aspects of new words.

In conclusion, although there does not seem to be an identified blanketing combination of direct and incidental learning, it does seem agreed upon that there seems to be a definite process that needs to be completed in order for a learner to acquire a new lexical item. When planning instruction for specific learners, these principles need to be met. Depending on the leaner though, one will find a different combination of incidental and direct learning.

3.2 Parts of Vocabulary that can be Taught

With some basic principles in place, it is now important to decide on what type of vocabulary should be taught. There are several suggestions given on what vocabulary can be taught. In an attempt to narrow down such a broad topic, one can consider items such as high frequency words, words parts, phrases, and learner goals.

Learners should be exposed to and learn a large number of words at the very beginning (Nation 44). Since there are over a million words in the English language, it can be a daunting task to decide which words are most important. Nation tells us that there are about 2000 high frequency words which comprises about 80% of the English language, 800 word families which is about 8%, 2000 technical words (4-5%) and 123,000 low frequency words (Nation; Bilingual Lexicon 124). Knowing that a learner needs to know about 90% of vocabulary for comprehension, it seems logical to begin instruction with the high frequency words followed by instruction pertaining to vocabulary specific to the learner's need. For example, if there is a learner who wishes to attend an institute of higher education there is an academic word list that is comprised of 570 word families (Nation 188). Academic vocabulary is a type of high-frequency vocabulary for those in academics, and it would be well justified to spend time on this type of word list. There are word lists for other specified areas such technology, business, agriculture, and many others. If one wishes, there are dictionaries that will separate words into semantic fields which can be used for instruction. There have even been lists created for high frequency collocations in academic writing (Hyland 51). So, it is possible to look at a word's frequency and range to determine an order of vocabulary instruction (Nation; Bilingual Lexicon 126).

Another approach to vocabulary instruction is to teach word parts. O'Rourke sites research done on memory patterns as support for a strong morphological component of vocabulary instruction (98). Teaching word parts can be helpful to the learner (O'Rourke 67). Knowing that a word is made up of certain affixes and stems and knowing the acceptable combinability of those affixes and stems can help the learner with decoding the words when listening, reading, and constructing words during speaking and writing (Nation 46).

The teaching word parts approach falls in line with Ellis's belief that SLA is accomplished primarily by the learning of sequenced language, ranging from the the phonemes to phrases (Ellis 101). Ellis states, "lexical phrases are as basic to second language acquisition as they are to first language acquisition, and so instruction relies as much on teaching useful stock phrases as it does on teaching vocabulary and grammar" (97). As discussed previously, language seems to be stored in chunks. If language speakers do store multi-itemed words the same way as single item words (which is one assertion of this paper), then there should be time dedicated to the instruction of set phrases in standard vocabulary instruction. As seen, although there is a wide choice of items to choose from there are ways to begin to narrow our choices.

Another way to narrow the options of vocabulary that could be taught is to create learning goals. O'Rourke feels that before beginning any type of vocabulary instruction, one first decides on the learning goals of the instruction (88). The goal will influence the type of vocabulary and how it is taught. If the goal is to increase the high frequency vocabulary, then one might choose a more direct approach or a more indirect approach (as noted above). If the goal is to teach low frequency words that are specific to certain subject areas, the instruction would focus on a smaller set of vocabulary items and may incorporate some semantic analysis. Identifying goals before instruction will help the leaner focus on a final outcome, and it will help the instructor stay focused on the appropriate materials.

Once the goals have been set, O'Rourke suggest finding out what items are currently in the learner's lexicon and move from the known words to the unknown (64). Moving from the known to the unknown will help students make connections between items already in their lexicon. The idea that previous knowledge of a learner can be useful in the acquisition of new

knowledge is seen in other areas. In Gathercole and Baddeley's research, they studied the

connection between the learner's ability to acquire words and the phonological complexity of the

word. They suggest the use of existing phonological vocabulary knowledge to learn the

phonological forms of new words (52). Also, one of the first steps of reading comprehension is to

have the learner activate his prior knowledge in order to make connections to the new

information in a text. This process helps the learner retain and understand the new material

(Mikulecky 41; Aebersold & Field 16).

These have been a few suggestions for deciding what type of vocabulary to teach. In

addition it is important to have a broader vision of what to teach in vocabulary instruction.

"Receptive knowledge enables students to comprehend word meanings appropriately. However,

to attain the ability to use the word fluently and with accuracy, ESL students have to acquire

more than just the denotative meaning" (Wei 1).

> It is not sufficient only to know the meanings of individual words. They need to be aware
>
> of the differences, the different skills that the language learners and users employ in the
>
> sentence production process, word perception, word formation, and word combinations,
>
> collocations and phraseology. (Shahhaidaripour 5)

Both of these quotes stress the importance of going beyond the traditional memorization of

wordlists with their single definitions or close synonyms. They speak to the need of having the

second language learner studying and knowing a word as an adaptive part of the language. The

learner needs to know that a word may have a different meaning in different written and verbal

situations, in combinations with other words, in different social situations, or geographical

regions.

In summary, the main focus of many ESL/EFL classrooms is to improve the communication abilities of their students. Since communication requires a productive vocabulary, methods and strategies should focus on how to move receptive vocabulary into the productive vocabulary. One major way to do this is to have meaningful output, which allows the learner to use new words in a variety of context (Nation 32). Kim suggests that vocabulary instruction needs to involve the production of target forms in order to effectively retain new vocabulary (286). This is not surprising, considering that the importance of meaningful output has been found to be effective in other areas of SLA such as grammar rules and formation of questions (Swain &Suzuki 564). Also, one can see the importance of output when looking at suggested steps for learning vocabulary. Quite often the final step involves having the learner use the word in a productive manner (writing or speaking) (Hatch &Brown 373; Nation 107).

Furthermore, vocabulary instruction should allow a learner to encounter new words in different contexts. It should give the leaner amply opportunity to apply and repeat new lexical items. Although incidental learning may allow the learner to acquire the highest number of lexical items, it is important to give instruction that has both breadth and depth. Also, the shear size and complexity of a language's vocabulary makes it impossible to teach all included items. It is important for the developer of a program to consider the intended audience, look at their goals and decide what will be most beneficial for the target audience. In doing this, the learners' time will best be used and allow them to focus on their specific needs. All these basic principles of vocabulary instruction need to be taken into consideration when developing an approach to vocabulary instruction.

3.3 Advanced Language Learners

Advanced language learners (ALL), like all levels of language learners, have struggles and

challenges that are unique to them. While a beginning language learner works on the acquisition

of basic vocabulary, phrases, and sentence structure, the emerging language learner pays more

attention to higher leveled vocabulary, more complex sentence structures, and finer grammatical

points. Even though the advanced language learner may have a firm command of the language's

grammar, syntax, and lexicon, he forever seems to struggle with sounding native. For a majority

of advanced language learner errors, the question is not whether or not the utterance is erroneous

or non-erroneous, but rather whether the utterances is acceptable, intelligible, or native like

(Lennon 189). As previously stated, ALLs do not struggle with the same problems as beginning

language learners or emerging language learners. For the most part they have a grasp on their

second language, so why is it important for the leaner to sound native-like and intelligible? The

answer is this, the ultimate goal of language is to effectively communicate. Even though the

learner may be able to construct non-erroneous language, if the message is being lost or

misconstrued, then there is a break down in the system and that needs to be addressed.

Researchers have identified certain problematic areas for ALLs. Two of these areas are in using

native sounding word combinations (collocations) and choosing words that fit the context both

denotatively and connotatively.

      Llach (2004) conducted research on advanced language learners at a Spanish university.

The focus of Llach's study was the effect of lexical errors on the quality of compositions. For

many ALLs, this is an important issue because their academic success is largely dependent on the

scores they receive on term papers and written exams, and as Howarth points out, a NNS may

fail to effectively communicate ideas because of incomplete linguistic competence not academic weakness (161).

All of the ALLs were taking part in an English business course and received the same prompt for their compositions. After the compositions were completed, Llach had two native Spanish speaking English instructors score the overall quality of the compositions and identify the number of lexical errors committed by the learners. Llach found a significant correlation between the number of lexical errors and the overall quality of the compositions (51). There was also evidence in the study that showed a correlation (although not significant) between the number of semantic lexical errors and syntactic lexical (collocation) errors. Llach's research provides two problematic areas for ALLs. The first is ALLs have difficulty using both connotatively and denotatively appropriate words in their writing; and the second is ALLs have difficulty with collocations.

These findings have been supported by various other researchers. Shahheidaripour states, "learners often create inappropriate nuances in the choice of lexical items because they are unaware of the extra senses that the words have" (2). At the end of the study, Shahheidaripour found that language learners struggle with knowing the nuances of words and therefore misuse them or collocate them with inappropriate words (11). More specifically related to collocations, Nesselhauf's (2003) study which was reviewed in the collocation section of the literature review, found collocations to be problematic for ALLs. She found that the highest rates of mistakes occurred with words that had a medium degree of restrictions (233). Fifty-one out of sixty-five of the collocational errors were the result of choosing an incorrect content word (Nesselhauf 232). Howarth echoes Nesselhauf's findings and states, "restricted collocations

occur much more than idioms and present a larger problem for learners" (169). In Bahns and

Eldaw's research, it was posited that advanced L2 learners have difficulty with word

combinations (101). Their findings showed that subjects produced more than twice as many

errors in their translations of collocations as in the translations of general lexical item.

In summary, the research has shown that ALLs struggle with collocations and fully

understanding the semantic range and appropriateness of some words. Although advanced

language learners have studied the target language for numerous years and have a strong

understanding of the grammatical and syntactical structure of the language; and although they

have developed an extensive vocabulary for the language, they still struggle with understanding

the finer colorings and nuances of the words. They struggle with sounding native-like because of

miscollocations, which are also likely the result of not knowing the the full semantic range of a

words. Sometimes, these errors effect communication and sometimes they don't. However,

although ALLs are able to successfully communicate a message, these errors can effect their

academic success, professional success, and ability to sound native.

3.4 Vocabulary Instruction for Advanced Language Learners

Although one goal of vocabulary instruction is to move words from the receptive to the

productive, caution ought to be held. When the move from receptive to productive happens, the

meaning and the form must be solidified. If not, the only thing that has been accomplished is that

the learner is now able to correctly produce a word with erroneous semanticity.

Following the Bahns and Eldaw findings, there does not seem to be a correlation between

the size of the learners' vocabulary and their ability to produce correct collocations. Therefore,

standard vocabulary instruction, in which the goal is to simply expand the learner's vocabulary is

not sufficient. Advanced learner vocabulary instruction should have a stronger focus on, not only the semanticity of the words, but also their collocations. This type of instruction also follows the assertion made by Robinson that to know a word is to know its collocations and how it is used in the language (275).

Regardless of whether or not a NNS is able to successfully paraphrase an unknown lexical item, the goal of language instruction should be to give the learner the correct lexical item for a given situation. It should be the goal of all speakers to effectively communicate an idea (Bahns and Eldaw 108).

Looking at the research from collocations makes a fairly strong argument for focusing on collocation instruction for the advanced learners. Also by looking at the research of short term memory and how it works, by looking at the research that was done on the use of restricted collocations, and by looking at they types of errors made by advanced learners, one can easily further support a focus on collocational and semantic range instruction.

4. Lexical Error Analysis of Advanced Language Learners

This study is similar in some ways to the reviewed studies. First the focus is on lexical collocations, with this being said, some of the identified collocational errors may contain errors in grammatical word choice, such as prepositions. So, although a misused preposition could be seen as a grammatical error, it falls within the limits of a miscollocation because the learner does not fully understand the word's collocational limits. Secondly, this study uses a written corpora. Written copora was chosen because writing allows learners to plan out their production in a low-pressure environment. This reduces the possibility that learners errors are a result of some outside stimuli such as the pressure often felt when verbally conversing with another person.

Furthermore, like Nesselhuaf, this is an exploratory study in terms of L1 collocations on semantic lexical errors. This is were this study deviates from the reviewed studies. To the knowledge of the researcher, there has not been any other study done that focuses on the relationship between semantic lexical errors and L1 collocations.

As we have seen, many of the terms have varied definitions. Therefore, for this study a 'lexical error' will be any error that is the result of any content word. A 'content word' is defined as a word that falls into a verb, noun, adjective, or adverb word class. A 'semantic lexical error' is a lexical error that is the result of a misdistribution of a word's semantical range. This can be the result of not knowing a word's connotational or collocational distribution. A 'direct translation' is a lexeme or lexical phrase that has a one-to-one word correlation with a parallel word order from the L1 to the L2. In terms of Lennon's error domain and extent, the linguistic unit of discourse sets the upper limit for the domain. In regards to extent, the lower level was the morpheme and the upper limit was the sentence.

Research Questions

1. What types of lexical errors are advanced learners committing in their writing?

2. Are the semantic lexical errors a result of L1 collocational knowledge in the form of direct translations or L1 claques?

3. Are the semantic lexical errors the result of a lack of knowledge of content words?

4. What are pedagogical methods to deal with the identified lexical errors, taking into consideration their sources?

4.1 Methods

4.1.1 Subjects

Subjects of this study consisted of two different groups. The first group consisted of NNS

sophomore and junior students enrolled in a university in the Ukraine. They were attending the

university to become English as a Foreign Language (EFL) instructors. The majority of the

participants' L1 were Russian, but there were some bilingual Russian/Crimean Tatar and

Russian/Ukrainian with Russian as their first language. All participants in the second group had

had English in their compulsory education beginning in the fifth grade, unless they attended

English schools, in which case, they would have received English instruction in the first grade.

The second group of subjects consisted of six Korean students ranging in age from 22-42

attending an American university as graduate students completing their last year of their Master's

degree in Teaching English to Speakers of Other Languages (TESOL). All students were

accepted on the basis of their TEOFL scores, but were still required to take an English

proficiency exam upon arrival. After the exam, it was suggested to all subjects that they take

further English instruction. Four of the six students had never studied in an English speaking

country. Of the other two students, one had spent three months studying in England, and the

other one lived in the Philippines for several years where she was instructed in English.

4.1.2 Materials

The corpus of essays consisted of 31 essays. Fifteen of the essays were Russian speakers

in the Ukraine and sixteen of the essays were collected from graduate Korean students. The

corpus consisted of 39,440 word tokens.

The Russian speaking students students were given the same prompts for the essays. There were four different prompts and they were designed to elicit nonfiction expository/ informative and persuasive responses. All of the essays were written in English, and there were no limits put on the length of the essays. The essays were collected in electronic format. The average Ukrainian essay was 1,198 words long with the longest being 2,513 and the shortest being 491 words long. There were a total of 17,973 words in the Ukrainian corpus.

The essays collected from the Koreans came from two sources. Twelve of the Korean essays were term papers from their first semester at the university. Although they did not have the same prompts, they were all on the same general topic. Since the twelve essays were term papers, they were heavily based on academic research and contained information from secondary sources. The other four essays were informative essays that were collected from a preservice class that all the Koreans attended before the beginning of their first academic school year. In this case they all received the same prompt and there was no research involved. All of the essays were collected in electronic format. The average length of the essays was 1,411 words long with the longest being 2,400, and the shortest being 294 words long. The Korean corpus consisted of 21,467 words.

4.1.3 Apparatus

Two different questionnaires were designed for this study. The first questionnaire was designed to look at the learners' use of L1 collocations in their L2 writing. There were a total of twenty-one prompts that informants responded to using a five-point Likert scale. The informants were to judge the likelihood of a given lexical phrase existing in their native language as a direct translation. There were three different types of prompts in the questionnaire, and they were

alternated throughout the questionnaire. Prompts 1, 4, 7, 10, 13, 16, and 19 were all lexical

phrases that were identified as collocational errors in the essays. These were included to see if

they were direct translations from L1. Prompts 2, 5, 8, 11, 14, 17, and 20 were all errors that

were marked as 'paraphrases' during the error analysis. These were included because many of

the paraphrases sounded unnatural, and the researcher felt that they may have been the result of

L1 direct translations. Finally, prompts 3, 6, 9, 12, 15, 17, and 21were the control. These were

prompts that were the corrected form of collocational errors in the learners' essays. This was

done in order to determine whether the learners' L1 was the source of the errors. If learners were

marking the prompts 1 as direct translations and prompts 3 as not direct translations, it could be

reasoned that L1 is the source or the errors. If the learners were marking both prompts 1 and 3 as

direct translations, or vice a versa, the question would be, "Why did the students make the error

if both the uncorrected forms and the corrected forms are rated the same?" All of the prompts

were given extra context in order to make sure the meaning was clear.

Also, following Shin and Nations suggestion, all the chosen collocations did not cross

immediate constituent boundaries. This means that a collocation did not cross from part of a XP

into another part of an XP. Take into consideration the following sentence. **The old man gave**

his most prized possession to his brother. It can be broken down into the following chunks:

- *The old man gave his most prized possession to his brother*

- *The old man*

- *old man*

- *gave his most prized possession to his brother*

- *gave his most prized possession*

- *most prized possession*

- *prized possession*

- *to his brother*

- *his brother*

A collocation would have to exist within one of these boundaries in order to be valid. So, *old man possession* is not an option for a collocation.

The second questionnaire was designed to see if a learners' L1 word covered the same semantic area of two L2 words. This a notion that Stockwel, Bowen and Marten labeled a 'split category' (qtd in Chang 287). A total of 40 prompts were taken from the lexical error analysis. If the a lexical error was classified as 'wrong word, right semantic area,' it was used in the questionnaire. The identified lexical error was placed next to a word that was deemed more appropriate by the researcher for the context in which the error was found. The informants just needed to circle 'yes' if there were separate words for each of the two word prompts, and 'no' if there was only one word that covered the two word prompts. Extra context was provided for some of the pairs in order to help the informant decide on a semantic field for possible compound words. For example, '*garden* work' and '*yard* work.' The word *work* was given to cover the possibility that there are separate words for 'garden' 'yard' 'garden work' and 'yard work.'

4.1.4 Procedure

There were two phases for this study. The first phase was the lexical error analysis, and the second phase was determining the likelihood of L1 interference in terms of collocations and calques. For phase one, each essay was read by a native English speaking (NES) graduate

student in a teaching English as a Second language (TESOL) program. As the graduate student read an essay, he would highlight anything that did not seem like an utterance that a NES would make. At this point in the process, there was no differentiation made between the type of error. Which means that the oddity could have been the result of a syntactical or lexical error. A syntactical error is a word order that is unnatural in the the target language. For example, in English the adjective precedes the noun it modifies, or in Spanish the adjective generally follows the noun it modifies. A syntactical error also shows up as having incorrect gender or number agreements between different constituents. A lexical error is the result of wrong word choice or word combinations. For example, inappropriately combining a verb and a preposition or using a word that denotatively fits, but because of its connotational meanings sounds awkward to the native speaker.   The latter part of the definition is important to this study since one of the focuses is in on collocations.

After the initial reading, each highlighted portion of text was analyzed. If the highlighted error was deemed to be the result of a syntactical error, it was disregarded and was not dealt with anymore. If it was determined that the highlighted error was the result of a lexical error, it was classified as one of the seven identified classifications. Words were thought to be of the same semantic area if they appeared as synonyms in a thesaurus. The following lexical error classification was used for this study (see figure 5):

Table 5
Lexical Error Classification for the Current Study

1. Totally Wrong Word	This could be the result of a calque or simply the wrong word

2. Wrong Word, Right Semantic Area	This takes into consideration word synonymy, connotative errors, and feature errors
3. Phonologically Related Words	This is similar to a misspelling, but it differs in that the learner's word is a word in the target language. *Right* is intended, but *rite* is produced
4. Misspelling	The word resembles a target language word, but does not follow the orthographic rules of the target language *Consumption* is intended, but *consumtion* is produced
5. Incorrect Collocation	Could result from a calque or from lack of L1 knowledge
6. Word Formation	Formed from L1 derivational rule or a misapplication of an L2 derivational rule
7. Paraphrase	Learner uses known words to describe an unknown lexical item

The second phase of the research was aimed at exploring the reasons for the semantic lexical errors. Semantic lexical errors were any lexical errors that were classified as (2)wrong word, right semantic area" or a (5) "incorrect collocation". For this portion of the research, two different questionnaires were developed and administered to informants of Russian and Korean first languages. The Russian informant was a faculty member of the TESOL department at an American midwest universtiy. His background is in applied linguistics, lexicography and TESOL. The Korean informants were three Korean graduate students in a TESOL program. One informant had been teaching English high school in Korea for approximately twenty years.

Another informant was an elementary English teacher in Korea, and the third informant had been

teaching English to businessmen in Korea.

All of the questionnaires were administered with the researcher present, and the

researcher was able to ask clarifying and probing questions; furthermore the informants were

also able to ask clarifying question. The informants had as much time as they needed and were

able to go back make any changes they deemed necessary. In respect to the Korean informants,

they were able to converse with each other and discuss their thoughts on specific items.

4.2 Results

4.2.1 Lexical Error Analysis

4.2.1.1 Ukrainian Essays

After performing a paired sample t-test, it is found that the number of lexical errors (M=12.46,

SD=7.19) is significant when compared to the number of total errors (M=21.93,

SD=13.73). Lexical errors made up 56.84% of all the errors made. Semantic lexical errors made

up 52.94% of all the lexical errors made. The most common error found was 'wrong word, right

semantic area. It made up 22.8% of all errors and 40.11% of all lexical errors. The second most

common error was shared by 'totally wrong word' and 'incorrect collocation.' Each of them

made up approximately 7% percent of the total errors and approximately 13% of all the lexical

errors (see figure 1).
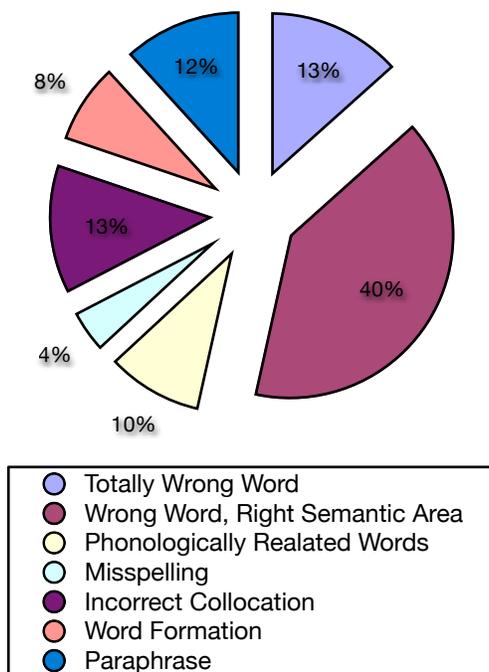
(Fig. 1. Ukrainian Lexical Error Distribution)

4.2.1.2 Korean Essays

After performing a paired sample t-test, it is found that the number of lexical errors (M=11.56,

SD=6.7) is significant when compared to the number of total errors (M=22.56, SD=16.19).

Lexical errors made up 51.25% of all the errors made. Semantic lexical errors made up 48.65%

of all the lexical errors made. The most common error found was 'wrong word, right semantic

area. It made up 17.45% of all errors made and 34.05% of all lexical errors made. The second

most common error was shared by 'totally wrong word' and 'paraphrase.' Each of them made up

approximately 9.5% percent of the total errors and approximately 18.5% of all the lexical errors
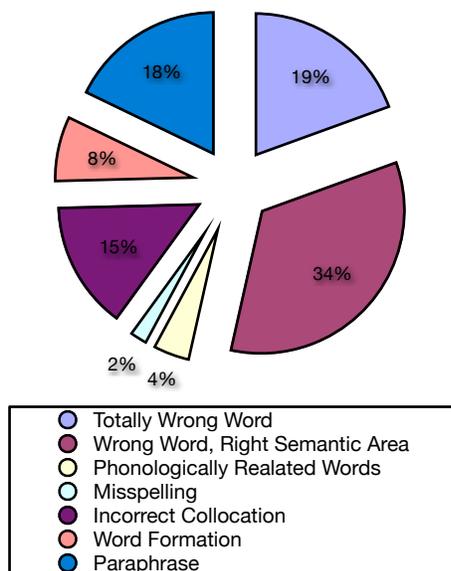
(see figure 2).

(Fig. 2. Korean Lexical Error Distribution)

4.2.1.3 Ukrainian and Korean

After performing a paired sample t-test, it is found that the number of lexical errors (M=12, SD=6.83) is significant when compared to the number of total errors (M=22.26, SD=14.48). Lexical errors made up 53.91% of all the errors made. Semantic lexical errors made up 50.81% of all the lexical errors made. The most common error found was 'wrong word, right semantic area. It made up 20% of all errors made and 37.10% of all lexical errors made. The second most common error was 'totally wrong word' making up 9% of total errors made and 16% of all lexical errors made. It was followed by 'paraphrases' and 'incorrect collocations.'

4.2.2 Questionnaires

4.2.2.1 Collocation Questionnaire

The questionnaire exploring the possibility of L1 direct translations as source of collocational errors yielded several interesting results. First, in terms of the Ukrainian essays, all of the 'incorrect collocations' (prompt type 1) were all rated as highly likely to occur as a direct translation in Russian. In contrast, all of the 'corrected collocations' (prompt type 3) were also

rated as highly likely to occur as a direct translation in Russian. In terms of the 'paraphrase' category three of the four prompts were rated as 'highly unlikely' to occur in Russian.

In regards to the questionnaire for the Korean Essay, all of the 'incorrect collocations' (prompt type 1) were rated as 'highly unlikely' or 'unlikely'. In contrast, all of the 'corrected collocations' (prompt type 3) were also rated as 'highly unlikely' or 'unlikely' to exist as a direct translation from Koren. There was only one item that was rated as 'highly likely' and that was a paraphrase.

4.2.2.2 Word Coverage Questionnaire

After analysis of the completed questionnaire, it was found that the Russian informant identified seven words that did not have separate words for the given prompts. One example is that there are not separate words for the words *relation* and *relationship*. The same is true for the words *wish* and *desire*. He identified four words that could have separate words in Russian, but could also be expressed by the same in Russian. For example, *decisive* and *deciding* do have different words in Russian, but there is enough semantic overlap that one Russian word could cover both of the English words. Finally, the majority of the prompts did have separate words for each of the provided English words.

With regards to the questionnaires the Korean informants completed, two of the informants said that one hundred percent of the prompts had separate words in Korean, while the other informant marked six sets of words that were covered by one Korean word. Two of the words that apparently do not have separate words are *aim* and *goal*. The reason for the discrepancy is unknown. Perhaps the other two Korean informants went outside of the semantic field for the word *aim* and where thinking of *aim* in terms of firing a weapon.

4.3 Discussion

4.3.1 Question 1: What types of lexical errors are advanced learners making?

Table 6
Percent of Lexical Errors of All Errors Made

	1	2	3	4	5	6	7
	Total Wrong Word	wrong word, right semantic area	Phonological Related Words	Misspelling	Incorrect Collocation	Word Formation	Paraphrase
Ukrainian	7.6%	22.8%	5.47%	2.43%	7.29%	4.56%	1.47%
Korean	9.97%	17.45%	2.22%	1.11%	7.48%	3.88%	9.14%
Combined	8.84%	20%	3.77%	1.74%	7.39%	4.2%	7.97%

Lexical errors have been shown to be a major error in the writings of both Ukrainian and Korean writings. Lexical errors contributed to almost one half of all the errors committed by advanced Ukrainian and Korean writers. This is quite significant considering that lexical errors have been deemed the most disruptive to comprehension in L2 writings (Dordrick 303).

A closer look at the data shows that the most common type of lexical error for both groups were "wrong word, right semantic area". This type of error could be attributed to an underdeveloped knowledge of vocabulary. The learners were using the words that were in the correct semantic field, but because of the connotative meaning of the words used, it did not fit the context. These errors are considered connotational errors.

There are two possible reasons for this. The first reason could be the learners do not have the same amount of words to cover the semantic field as there are in the L2. This is what study question three investigates. The second plausible reason is that the learners do not have a full knowledge of the word. This would mean that they do not know the additional connotative

meanings (formal, informal, slang, etc.) or the appropriate collocates. An important part of knowing a word is when, where, and how to use it.

The second most common lexical error committed by the two groups was classified as "totally wrong words.' The reasons for this error could be numerous. The writers may have accessed the wrong word in their mental lexicons because of how it is being stored. Perhaps the word belongs to their receptive vocabulary; therefore, there is not easy access to it for production. Perhaps the target word and the produced word were phonologically similar, and as a result, the learners thought they were producing the target word.

The third most common error was 'paraphrase.' A moment should be taken to explain that although the paraphrase was marked as an error in this study because a target word was not produced, in general, paraphrasing is a useful tool for learners to use (Zimmerman 40). It perhaps is more useful in spoken language since spoken language happens in real time and does not allow for planning. In writing however, since it is a planned activity, time should be taken to identify the target word rather than paraphrasing it.

The fourth most common error was 'incorrect collocation' and it happened only four less percent of the time than the 'paraphrase.' There may be a few reasons for miscollocations. The first may be that there is L1 interference and the produced collocation is the result of a direct translation. This is the focus of research question two and will be discussed later. Another possible reason is an underdeveloped knowledge of the word. The word may have been in their productive vocabulary, but the learners may not have understood the different connotational colorings, resulting in putting the word in the wrong context. Also, there was a mix between lexical collocations and grammatical collocations. Grammatical collocations such as *refer to*

were often produced incorrectly, omitting the prepositions. These types of collocations could happen if the leaner struggles with use of prepositions.

Many of the collocations fell into the category of restricted combinations; although there was at least one instance where an idiom was produced incorrectly. The target idiom was *opposites attract* and *opposites are attracting each other* was produced. A further, more in-depth discussion about the collocations will occur during the discussion of study question two.

There were few 'word formation' errors. They only accounted for 4.2% of the overall errors and 7.4% of the lexical errors. These types of errors often occur for two reasons. The first is the misapplication of an L2 derivational rule, or sometimes, the leaner applies an L1 derivational rule to the production of an L2 target word. This is beyond the scope of this study, but would be an interesting research topic for a different study.

There were some minor discrepancies between the error distribution between the two language groups. Although the main type of lexical error was the 'wrong word, right semantic area, 23% of all errors were classified as type 2 for the Ukrainians and only 17% for the Koreans. This discrepancy is most likely due to the level of the students. The Korean students were graduate students in the last year of their degree, while the Ukrainian students were sophomores and juniors in an undergraduate degree. The types of language are not believed to make a difference in this discrepancy because this type of error has to do with word knowledge and not word order or other grammatical issues.

There was quite a difference between the two groups concerning paraphrasing. The Korean writers (9.14%) seemed to use more paraphrasing than Ukrainian writers (1.47%). This could also be the result of a more advanced level. Paraphrasing is a tool used by advanced

learners to navigate communication. One more discrepancy to mention is between the 'derivational errors.' It is possible that the Ukrainian writers made more of these types of errors because their language is highly inflected and regularly uses inflections to mark word class.

In summary, half of all errors made were classified as lexical errors. The most common types of lexical errors that were found had to deal with having a limited understanding of the semantic range of the words and how they collocate with other words. These types of errors accounted for 53% of Ukrainian lexical errors and 48% of Korean lexical errors.

4.3.2 Question 2: Are the semantic lexical errors a result of L1 collocational knowledge in the form of direct translations?

In view of the current research, it seems that most of the errors that were identified as collocation errors are not the result of L1 interference. There are three main reasons for this. The first is that both prompts one and three from the survey received the same ratings from both of the language groups. In the Ukrainian language group, both of the uncorrected and corrected collocations were deemed to be 'highly likely' to occur as direct translations. This shows that learners were not using direct translation in their writing. If they were using direct translation, then prompt type three would have been marked as not being possible as a direct translation. If the prompt type three was able to be directly translated, then the error would have never been committed in the first place. For example, a learner produced The ***interest to*** old Russian *paintings*. The grammatical collocation *interest to* is unacceptable. In the questionnaire, the corrected form of this grammatical collocation was provided as The ***interest in*** old Russian *paintings* and was scored as 'highly likely' to exist as a direct translation. If *interest in* is 'highly

likely' then why wasn't it produced? The most probable answer is not that it was a direct translation, but rather the learner does not know the correct L2 collocation. It could be argued that this example shows that the learner does not know the use of prepositions, but in cases like this, the preposition *in* has a more semantically opaque meaning. A preposition like *of, about,* or *on* would fit better with idea that *interest in* conveys.

The assertion that the miscollocations were not the result of direct translation is further proven by the Korean questionnaires. All of the incorrect collocations and corrected collocations were rated as either 'highly unlikely' or 'unlikely' as existing as a direct translation. This exemplifies the same point that was made above. If direct translation was a valid reason for the miscollocations, then the prompt type 1 should have been marked as being highly likely to exist as a direct translation in the L1.

An interesting finding from the questionnaire concerns prompt type two in which the paraphrases were rated as likely being direct translations of the L1. Interestingly, these were the only types of prompts that were thought to be "highly likely' in both language groups. Although this was not a research question of this study, it may lend itself to further study.

In light of the results found in this study, there is still a strong belief that collocations are an important part of L1 acquisition and SLA. It would be interesting to do further research to look at some of the other general errors with collocations in mind. During the analysis of the essays, it was noticed that many of the highlighted oddities that did not qualify as lexical errors were the result of nonnative like word order. This is reminiscent of Howarth's comment that there are many grammatical ways to communicate an idea, but there are a limited number of ways that sound native (162).

As stated above, almost all of the identified collocation errors fell into the category of restricted combinations. There were only a couple of exceptions, which were incorrect idioms. The discovery that most of the trouble was found with restricted collocations is in line with the focus and findings of previous research (Bahns & Eldaw 103; Chang 285; Nesselhauf 230; Howarth 163; Wei 2). Others like Hyland have studied 'academic clusters,' the focus is not the incorrect production or the items, but rather how their use can help with the cohesiveness of a paper (42). The reason for the main focus being on restricted combinations is because they are harder to acquire. Lexical phrases such as idioms and other fixed phrases have a fairly set lexical content and a set syntactical order. Students can approach such lexical phrases as a single unit. However, with restricted collocations, there is the factor of arbitrariness that was discussed previously. The lexical content of a restricted collocation is not as concrete and therefore, provides room for error.

4.3.3 Question 3: Are the semantic lexical errors the result of a lack of knowledge of produced words?

From the data that was found in this study, the answer is yes. In the view of the researcher there were three possible reasons why a learner would produce a word in the right semantic area, but have it judged incorrect considering the context. One reason would be that the leaner had an insufficient or incomplete knowledge of the word. The second reason might be that the learner has an L1 word that he wishes to use and does a direct translation of the target word with the collocates that are appropriate for that word in the L1. However, the target word has a different range of collocates in the L2 and L1, resulting in the unacceptable form. The third reason would

be the notion of a 'split category' in which two words in one language are covered by one word

in another language.

According to the data of this study, reason two seems to be invalid. Accordingly, reason

three seems invalid also. The Russian speaking informant indicated that 73% of the lexical errors

that were considered to be in the right semantic area had separate words for the produced word

and the acceptable word. Furthermore, two of the three Korean informants indicated that 100%

of the lexical errors marked as right semantic area had separate words for the produced word and

the acceptable word. This data clearly shows that although reason two does contribute to a few of

the semantic lexical errors, the majority of the semantic lexical errors are not the result of the

'split category' notion. With both two and three disproved by the information gathered and

analyzed in this study, it is assumed that a majority of the semantic lexical errors are the result of

the learners having an underdeveloped or inadequate knowledge of the words.

4.3.4 Question 4:What are pedagogical methods to deal with the identified lexical errors,

taking into consideration their sources?

As this study has shown, lexical errors make up a significant part of all errors made. It

has also been shown that semantic lexical errors make up 50% of all lexical errors made.

Through the discussion, it was shown that there are several possible explanations for the

semantic lexical errors, but the most likely explanation is that the learners have an

underdeveloped and inadequate knowledge of the words being produced or the target word.

Since this is the most feasible explanation in light of the current study, the pedagogical

implications will be aimed at improving the learners' knowledge of lexical items, and more

specifically clearing up the confusion of words in the same semantic field. As part of this process

activities will also focus on words and how they collocate with others, since the semanticity of a word and how it co-occurs with other words seem to be closely connected. For example, learners know the full semantic range of *field* as it relates to an area, they would know that it collocates with areas of study such as *the field of education, the field of linguistics,* or *the field of science*. If they had this rich understanding of the verb *field* and understood its multiple denotative meanings and its full semantic range, perhaps they would be less likely to use it as *the field of schools*.

The first step in suggesting learning strategies is focusing in on what parts of words the learners do not know. This idea is partially knowing if word is valid. Simply consider the passive and active vocabulary. If a word is in the passive vocabulary, it can be recognized, but not produced. A word in the active vocabulary can be recognized and produced. It would not be said that the learner does not know the word in the active vocabulary, but it could be said that the learner does not know the word in the passive vocabulary as well as he knows the word in the active vocabulary (Baumann & Kameenui 606). Along with the idea of passive/receptive and active/productive vocabulary, it should be noted that there in not a clear defining boundary between the two types of vocabulary. Rather it should be seen as along a continuum. This is true for both native speakers and non-native speakers of a language. People may have words on several parts along the continuum. For example, a word may be on the far end of the continuum, where the learner has heard the word before, but has no sense of what it means. A little further down the continuum, a learner might recognize a word and have some sense of the meaning. Perhaps the leaner recognizes it as a word used in economics or biology, but he cannot articulate any sort of definition. Continuing down the continuum there are words that the leaner recognizes

and can use, but does not know the full meaning of the word and will use it incorrectly in certain contexts. Finally at the far end of the productive vocabulary, the learner has a full understanding of the word. He can articulate its multiple denotative meanings, comment on its connotative meanings, and describe what words it can successfully collocate. It seems that most of the errors committed by the ALLs in this study, fall into the third category of having some sense of the word, but not fully understanding its denotative and connotative meanings. Nor do they understand with what words it can successfully collocate.

As was mentioned in the beginning of the paper there are many aspects to knowing a word. It is obvious that if the learner is able to almost successfully produce the word, then there must be certain aspects that the learner knows. Nation's criteria to knowing a word can provide us with a basis for determining which elements of the word need to be focused on. To review these are Nation's criteria for knowing a word.

1. spoken form

2. written form

3. word parts

4. connecting form and meaning

5. concepts and referents

6. associations

7. grammatical functions

8. collocations

9. constraints on use

The data for this study will not be able to comment on some of these criteria. For example, there is no way of knowing whether or not the spoken form is known because this study used essays as a language medium. Furthermore, this study cannot determine whether or not the word parts or associations are known. A different type of assessment would be needed to reveal this information. However, this study does show us that the written form is known as is the connection between form and meaning. It is believed that form and meaning are known because the learners produced a word in the correct semantic area, which would be the concept. The errors classified as "totally wrong word' would be more appropriate for stating that the learner has not connected the form and meaning. Similarly, it is thought that the learners have to know the concepts and referents of the words identified as errors. This is thought for the same reasons as stated in regards to 'connecting form and meaning.' This study will not concern itself with these criteria because they do not apply to the semantic lexical errors that are under discussion. It is also agreed that the learners understood the grammatical functions of the words that were marked as errors. If a marked lexical error was used in an inappropriate word class, it would have most likely been classified as a 'derivational error.'

The criteria that are thought to be underdeveloped in regards to the production of the semantic lexical errors are 'collocations' and 'constraint on use'. It is felt that knowing a word's collocates and its constraints on use are closely related. If a learner knows that certain a word comes from the formal register, then he would know what words would properly surround it, and vice a versa. If the leaner knows what other words regularly surround a word, it may give clues to its constraints. However, it is also thought that although these two areas overlap they are distinctly different. Just because a learner knows that *bitterly* collocates with *cold* it does not

necessarily mean that the same learner can make a judgment on formality, frequency, or currency.

5. Pedagogical Suggestions for Advanced Language Learners

In review, this section will focus on different methods, strategies, and considerations for vocabulary instruction with a focus on improving the advanced language learners' knowledge of a word's collocates and usage constraints. This focus has been chosen because it was found that lexical errors are a significant problem in advanced learners' writing. In addition, almost fifty percent of the lexical errors were semantically related, and collocational knowledge and usage constraints focus on the semantic aspects of knowing a word. The strategies and activities will incorporate many aspects which were discussed previously in the "vocabulary instruction" portion of the literature review. Some activities will take an explicit and direct approach in which the learner will directly look at the attributes of words and collocations. Some activities will incorporate the skills of reading and writing to give the learner multiple opportunities to encounter the word. Also, this section will include concrete examples so that the reader will be able to easily transfer the findings and ideas of this study and apply them to the classroom.

5.1 Teaching Collocations

Both this study and several others have asserted the necessity for teaching collocations (Nesselhauf 223, Bahns &Eldaw 109). Teaching collocations will meet the needs of advanced learners. It will help their language production seem more native-like and increase their understanding of the semantic range words. In choosing activities and strategies for collocations, several things were kept in mind. Wei points out that collocational competence is reached by building awareness of how words work in combination (8). One of the main goals of this type of

instruction is for the learner to sound more native-like and therefore become more intelligible. Shin and Nation believe teaching collocations is an efficient way to do this (340). As noted before, the restricted collocations create the biggest problems for learners. Because of this, it is important to identify what collocations need to be taught in order to focus specifically on those (Hyland 44). There is a concern that the over stress on phraseological knowledge can impede the learner's creativity skills (Granger 158). The activities presented will try allow enough freedom so the creativity of language is not stifled. Finally, the success of the vocabulary instruction will be dependent on the amount of involvement of the activity (Kim 311). The activities suggested will keep this in mind in order to optimize the time and effort of the learner and the instructor.

5.2 Considering the Learner

Each learner has different needs. In order to meet the specific needs of the learner, the instructor will need to take the student's current knowledge and goals in to mind in order to best help the learner. This process may look different depending on the type of course one is teaching. At a minimum, the instructor can choose the type of vocabulary to focus on taking into consideration the focus of the course. If one is teaching an English composition class to exchange students at a university, the focus of the collocations might be on common collocations used in academic writing. If one is teaching an English course to NNS MBA students, the focus of vocabulary might be on commonly occurring words in the business and academic world. There are several lists of high frequency collocations available. Hyland (2008) provides a collection of the most frequent four word collocations that was derived from the British National Corpus Baby edition. In the same article, Hyland provides a list of the most frequently used four-word

collocations used in research papers, PhD theses, and master's dissertations (51). These types of lists can be used to help guide instruction

If more time allows, it is a good idea to come up with goals either as a large group, in small groups, or individually. This type of approach works best if the class has more freedom in the content that is covered. For example, a conversational class that a university or city provides to people as service will allow the instructor more freedom. In this setting, one can begin the course by administering a survey to the students to identify the needs and goals of the students. The survey might include questions about the students' interests and their backgrounds. It can even include specific questions about what the learner would like to learn from the class or it can ask the learner what he struggles with most in the target language. This type of survey will serve two-fold. First it will help the instructor decide on specific vocabulary items to focus on, and it can provide a writing sample to gauge the level of the students in the class.This type of survey will help the instructor narrow down what type of vocabulary the groups, groups, or individuals will most benefit from. This type of approach also helps add to the involvement level of the learner.

Although the opportunity for individual meetings with students is not as common, when it is an option, it is a good idea to set individual goals with the students. If possible, the instructor can meet individually with the students and ask many of the same questions as stated above. Previous to the goal setting meeting, it is helpful to have the students produce a writing sample so that the instructor has an idea of the learner's ability. Furthermore, it is helpful in the initial meeting with the student to make some anecdotal notes on the student's level of speaking ability. Both of these items will help the instructor guide the goal setting process.

Whichever setting an instructor finds himself in, it is important for the goals that the group, small groups, or individual make to be attainable with the amount of time and resources available. Once goals have been set, it is important to check in periodically to see how students are progressing towards their goals.

5.3 Collocation Activities

After each activity, the advantages and disadvantages of each activity will be discussed. It is important to note the limitations of each activity because no single activity will fully meet the needs of the learner. By identifying the strengths and weakness of an activity, one should be able to successfully combine several activities and optimize the learning potential of each, which in the end, will benefit the learner the most.

Nation (2001) presents several activities that can help with chunking and collocations (337-344). The first set of activities involve the receptive skills of reading and listening. Using receptive skills, especially reading, has been a well accepted method for learning vocabulary. Although Nation agrees that large amounts of extensive reading can improve the breadth of a learner's vocabulary, it does not do much for the depth of understanding (148). He therefore suggests that there is a switch from extensive reading to intensive reading (151). In the intensive reading, the learner is more involved in the text he is reading. This also follows Kim's involvement load hypothesis (2008). For this reason, the following activities which incorporate reading, should be done with text that is familiar to the reader and at their independent reading level. Text at a learner's independent reading level is a text which the learner knows at least 99% of the words on a page. It is important that the learner is using text at the independent level

because he needs to be able to obtain the meaning of the text. If the learner spends too much time decoding the text, the comprehension is affected.

5.3.1 Listening Corner

This technique is used both with first language acquisition and second language acquisition. It allows the language learner to hear the language spoken by a native speaker (most generally the teacher) and then try to produce what he has heard. Depending on the setting of the classroom and the age of the learner this may or not be appropriate for the classroom setting. If the classroom setting does not lend itself to this activity or the leaners is at an age where they are not comfortable doing this around others, the activity can be done individually in a setting that the learner finds comfortable such as a study room in the library or even at home. With many of the new technologies available, this type of activity can easily be completed independently in a safe setting.

Before the learner can do the activity, the instructor needs to choose a piece of text that is familiar to the the leaner. The text can be fiction or nonfiction. There are benefits to both types of literature. The fictional text will be easier to understand because it has a story line that will help the reader with the comprehension; therefore, more cognitive energy can be spent focusing on the fluency of the text. However, nonfiction text, such as the learner's text books, would provide the learner with vocabulary that is specific to the area of study.

Once the text is chosen, the instructor makes a recording of the text. As the instructor reads the text, he reads it as naturally as possible so that the natural rhythm and cadence of the target language emerges. Depending on the age of the learner, one might be able to find books on tape that are appropriate for the learner. The listener should be able to identify the natural

chunking of words that happens in the language. After the learner has listened to the recording several times, there are a couple of options.

First, provide the learner with text that was on the the recording and a device for her to record their voice. Have the learner reread the text out loud and record it. After the students has recorded his voice, he listens to the original recording and compares it to the one that was just made. The learner's goal is to have his fluency sound similar to the instructor's recording. It is important to work with the learner beforehand and stress that the purpose of the activity is to have similar fluency, similar chunking of words. The goal of the activity is not pronunciation. It is important to highlight and isolate the skill that is being practiced.

The second option is for learners that might have difficulty with the first option that was presented. Therefore, another option is to once again provide the leaner with a written form of the text. As the learner listens to the recording, she marks where the natural breaks occur. These breaks will follow natural constituent boundaries of the language. After the learner has segmented the text, she can check her answers with a control copy of the text. After she has mastered this portion of the exercise she can go on and attempt a reading of the text. At this point the goal is to read the text with the same fluency as the native reader.

" When I was a young boy/, I lived a small town/ whose limits could be reached/ by walking twenty minutes in any given direction/. At the edge of our small town/, was the beginning of farm fields and wooded valleys/ as far as the eye could see/. It was in these fields and woods/ where I spent most of my days as a young child/. The woods and fields/ seemed to continually call to me/. As I stood on the edge of houses and homes/, there was

a continual calling/, urging me to enter/. Urging me to walk/ until the rooftops and

chimneys/ were lost to the leaves."

This activity benefits the learner in several ways. First of all, because the text is familiar,

it allows him to focus on the chunking that occurs on the recording. Since the reader on the

recording is ideally a native speaker of the target language, the learner will be exposed to the

natural chunking that happens at the phrasal level in the target language. This type of activity

also allows the leaner to be exposed to the same text multiple times if he needs it. As mentioned

before, a learner needs to be exposed to a lexical item several different times in several different

ways. In this activity the learner gets multiple exposure in reading, writing, and speaking. Also

with current technology, this type of exercise could be accessible via the internet, or at a

minimum, easily transported and accessed.

This activity does have its limitations. First of all, this activity focuses on the use of

implicit instruction. Although the leaner is explicitly being taught fluency and attention is being

brought to the natural chunking that occurs in a language, it is relying on incidental learning

when it comes to the learning of collocations. As seen in the the example above, the natural

breaks occur within constituent boundaries; however, there are collocational relationships within

those boundaries that are not being isolated. Secondly, this activity is not addressing the

connotational meanings of the words in the collocation, which is one major problem ALLs have

with their language production. ALLs do not know the full semantical range of a word and

therefore pair them with unacceptable or unnatural sounding words. Finally, this activity relies

heavily on the instructor. It requires the instructor to have a native-like control of the language. It

also requires the instructor to spend a considerable amount of time finding and preparing the

level appropriate materials. Both of these requirements could be a challenge depending on the location and resources of the instructor.

5.3.2 Collocations in Student Writing

The following activity is designed to be a formative process that could easily be integrated into the writing process of class assignment. The first thing the instructor will need to do is assign the learner a writing prompt. The writing prompt will depend on the learner and the academic setting. If the setting is specific, a course such as an English literature course, then the writing prompt should be specific to that setting. If there is more freedom in the setting, the student can choose the topic of the paper. Whatever the prompt is, the composition should be something that can be worked on over a period of time. Ideally, writing should be substantial in length. The more writing the learner produces the greater the possibility of the learner creating useful material for analyzing.

Once the initial draft is complete, the instructor goes through the regular writing process. Ultimately the instructor will go through the paper and look for misallocations in the learner's writing the instructor should be aware of the number of oddities she is noting. She does not want to overwhelm or discourage the learner once the composition is returned. An acceptable number of oddities would be between three and five per page. However, before this happens the learner's writing should be as developed as possible. It should be after several drafts as far into the writing process as possible. The leaner should have had several opportunities to review the overall structure of the paper and correct any spelling, grammatical, and punctuation mistakes. The one thing the leaner should not do is have a native speaker correct mistakes or make word choice

decisions for the learner. The instructor is going to want a paper that represents the learner's actual knowledge of the target language.

Once the instructor feels he has a true representation of the students mental grammar, the instructor will go through the learner's text and identify items that are not native-like or that distort the intended meaning. The instructor is looking for lexical items that denotatively fit the context, but not connotatively. Depending on what the instructor finds at this point in the analysis, two things could happen. First, if the instructor is finding errors that involve oddities that include verb-prepositions combinations, misused idioms, common adverbial phrases, or common transitional phrases, the instructor can take a more explicit approach. The instructor can use resources such as prefabricated list that include the more fixed collocations of the language. For example, if the student is misusing or incorrectly producing transitional phrases a list such as list 4.1 could be used.

Table 7
Transitional Phrase List

Transitional Phrases
In the beginning
In the first place
To begin with
In regards to
In addition to
With this in mind
At the same time
One can also consider
Another point of view
Contrary to
In the end
To summarize

In this case, the student can be given the list and encouraged to use the transitional phrases in his writing. When there are instances of students misusing highly fixed collocations such as these where a content word is placed with closed category word (prepositions, article, etc), memorization is an acceptable strategy.

However, if the oddities consist of restricted collocation (verb-noun, noun-adjective) combinations, then a different approach should be taken. In these instances, the learner most likely paired these words as a result of not knowing the full semantic range of one or both of the words. At this point, it is worth the learner's time to look closely at the lexical items. One way to to do this is to explore the semantic range of a word. The following chart could be used to analyze the word target word and its synonyms. Although several classifications have been given, many more exist.

Table 8
Semantic Range Analysis Example

	Word	Synonym	Synonym	Synonym
	Man	Gentleman	Guy	Dude
Formal Register		yes		
Informal Register	yes		yes	yes
Slang			yes	yes
Positive Sense	yes	yes	maybe	Yes
Neutral Sense	yes			
vulgar sense				
Negative Sense				maybe

	Word	**Synonym**	**Synonym**	**Synonym**
Used with Friends	yes	yes	yes	yes
Used with Family	yes	yes	yes	maybe
Used at Work	yes	yes	yes	
taboo				
politically correct	yes	yes	yes	
complimentary	yes	yes		

This type of exercise will help increase the learner's knowledge of the connotative meanings of words. With this additional knowledge, the learner would be able to make better judgements about how to pair words. For example, after this connotative analysis, hopefully a learner would not produce *"My boss is a very respectful dude*." Knowing that the word *dude* is an informal word teetering on the edge of being slang with some negative coloring to it, the learner would not pair *dude* with *respectful* when talking about the boss.

An extension for both of these options is to have the learner write sentences in order to experiment with acceptable collocations. If one does choose to do this extension, a native speaker of the target language would need to check the acceptability of the leaner's collocation.

When considering the benefits of these activities, let us first consider the use of the student's own original writing. There are several benefits to using the learner's writing as an basis for instruction. First of all, the vocabulary work that the learner is doing is very personalized. It is an accurate representation of the learner's mental lexicon; therefore, it can be

used to target the specific needs of the learner. It also increases the learner's involvement in the learning. The learner has ownership and vested interests in his own work. When it is the property of the learner, it can seem less like an arbitrary exercise and more like a learning experience. Finally, this is an example of finding out what is currently in the learner's lexicon and then moving forward from there. This follows O'Rourke's suggestion that moving from the known to the unknown will help the learner make connections between items already in their lexicon (64)

There are also benefits to using lists with some collocations. The reality is when one learns highly restricted set phrases such as idioms or the phrases we looked at above, memorization is probably the best approach. It is hard to semantically analyze these phrases because they are often opaque in meaning or they consist of several grammatical words (which tend to carry little meaning). Because of this, it makes it difficult for the the learner to make an informed judgement on what words to pair. Therefore memorization of these types of lexical phrases is necessary.

Analyzing words and their synonyms provides several benefits to the leaner. The leaner obtains an in depth understanding of a word. This type of analysis allows the learner to look at a word from multiple perspectives, and it gives the learner multiple exposures to the word. This is also the type of intensive word study that is necessary for a NNS to fully grasp the full range a word has in the target language.

However, this type of activity takes a lot of time. It is time intensive for both the learner and the instructor. It will take the learner considerable time to write a composition that will be analyzed. It will also take the instructor considerable time to correct the composition and provide meaningful feedback. One possible way to reduce the amount of time spent by both parties is to

incorporate this into an existing assignment. It could easily be a couple of extra steps within the already used writing process. A trade off the instructor may have to make is spending more time on a few pieces of writing rather than spending little time on multiple pieces of writing.

Although the memorization of some collocations is acceptable, it should not be the norm. Memorization may provide learners with chunks of usable language, but it does very little to help them gain an understanding of a word. A rich understanding of a word is a benefit to learners because it would allows them to use that knowledge when working with less restricted collocations. Finally, in regards to the analysis of the specific words and their synonyms, this process is also time consuming. Learners will most likely need access to a native speaker of the target language. Also, the sheer number words in a language's lexicon makes this process extremely inefficient; however, with narrowing of words through student goals and the setting in which the instruction is taking place, learners should be able to focus on the necessary lexical items.

5.3.3 Matching Words to Create Collocations

In this activity, the learner is provided with two list of words and its focus is on restricted collocations. In preparation for this activity, the instructor can narrow the words by means already discussed. This activity is best done in a large or small group because it will allow several people to weigh in on the acceptability of a word.
An example of a list might look something like this

Table 9

Example Word List

Column A	Column B
ride drive mount steer operate pilot handle	bike car train conversation bus machinery

The learner then takes the first word from column A and decides which of the combinations are acceptable. The learner would right out *ride a bike, ride a car, ride a train, ride a conversation, ride a bus,* and *ride machinery.* Once the list is written, the learner/s will discuss which ones they think are acceptable. After the discussion, the following might remain: *ride a bike, ride a train,* and *ride a bus*. If this is done in a group setting, someone might offer up *ride in a car*, which would be acceptable; however, *ride in a car* contains the phrasal verb *ride in* which could be deemed a separate verb than *ride*. Nevertheless, this type of conversation is healthy for the students and can lead them to a richer understanding of the words. The ultimate decision can be made by the instructor if the students wish to have a definitive answer.

As an extension of this activity, the learners can once again do some writing and include some of the new collocations they have created. This type of activity could also be coupled with reading that is being done in class. If there is a set of vocabulary words that is being presented with a given chapter of a book, the instructor could find synonyms for some of the words and have the students complete the activity.

This type of activity has many of the same benefits as the previous activity in which the learner is analyzing the semantic range and connotative meanings of words. The activity achieves the same goals in a more implicit way. This activity would be a good lead up to a world analysis activity similar to the one presented above. The activity also provides the learner with multiple exposures of a word. If this is done in tandem with literature that is being used in class, the leaner would be exposed to the lexical item in the text, in the activity, and once again in the writing extension.

The downfall of this activity is once again the time it takes to complete and the number of words that could be studied. In order for a learner to work through all of the combinations and then write a sentences which include the new collocations, the learner could easily spend an hour completing this task. The net benefit of the time spent would only be relatively small. This would also be a challenging activity to do alone or without the help of a native speaker to help guide the learner.

5.3.4 Concordances

A concordance is a list of contexts exemplifying a word or word family (Nation 111). Currently, concordances are created by gathering large amounts of text and uploading them into a large database. There are concordances of Shakespeare's works and of the Bible. There is the American National Corpus and the British National concordance. Concordances are made up both written and spoken language, so the user can get an authentic look at how a word is being used and in what context. Even with today's home computers, one has a concordance at his finger tips. Here is a concordance of the word concordance from this document which was created by searching for a specific word.

...and **concordances**. Melcuk found that...

...students use **concordances** to look at the context...

...of a **concordance** and collocations...

... a **concordance** is a list of contexts...

The use of a concordance can be very helpful to learners. It allows the learner to access authentic language and see how it is being used.

There is a sampling of the British National Corpus (BNC) that one can use online. There is limited access to the corpus without purchasing a subscription. However, one can still use the free online version, search for words or phrases, and see up to fifty results. When the results are displayed the word or phrase that was searched for is not highlighted, requiring the user to search a bit harder. With the free version, there are not options to limit or narrow one's search.

Mark Davies of Brigham Young University (BYU) has created a very useful online concordance called Corpus of Contemporary American English (COCA). Its many features make it ideal for a wide range of people from researchers to language learners. It creates its concordances from over 425 million words. The user can search for specific words or phrases. One is able to get results on the most common adjectives used with a noun. For example, if one wants to know the most common adjective used with car, she can put that information into the search engine and find out that *new* is most commonly used with *car*. The system also gives the user choices on how she wants the results sorted. It can sort by relevance (which is useful if one is looking for a specific word combination) or by frequency (how often the words co-occur). One can also limit her search by the type of media she wants to search. The system, by default,

draws corpora from spoken, newspapers, magazines, academics and many more. In order to get more specific result, the learner can choose one or several of these criteria for their search.

For this paper, the researcher also attempted to review the International Corpus of English and the American National Corpus. However, when it was attempted to access both of these corpora there were issues with downloading the data and gaining access to the programs. Since one goal of this portion of the paper is to create resources that are easily accessible to the reader, there will be no further discussion on these two corpora.

Concordances can be helpful tools for learners, but first some time needs to be spent on how to use the corpus. After looking at the three discussed corpora, the most convenient is the one made available by BYU. It is suggested that time is taken to teach students how to use COCA since there are so many different functions and options. The functions that would be most useful for advanced language learners would be searching for words and phrases, searching for specific collocates, and searching for collocates by word classifications (i.e. adjectives near a certain noun or prepositions near a certain verb). Once the learner has a firm grasp on how to use the corpus, there are several ways for the learner to use it.

Like other suggested activities where the learner is explicitly working with vocabulary, there needs to be some narrowing of vocabulary. Once again, depending on the the classroom setting or course type, the instructor may be able to provide the learner with a list of vocabulary that is commonly used in the area of study, or that students traditionally struggle with. Also like before, one option is to use an original piece of student writing and analyze it for collocational mistakes. Once this is done, the learner can then use the corpus to get a clearer picture of what words naturally collocate together. For example, let's say the learner incorrectly collocates

clench with *ears* in an attempt to convey the meaning of not listening. The learner would go to COCA put *clench* in the word search box, in the collocation box choose collocates with nouns within a two word range on either side of *clench*. The results will show that the *clench* and *ear* do not collocate. It will show that *clench* is most often used with *fists, teeth, jaw, stomach, hands, and fist*.

Concordances could also be used as way to check the validity of collocations while doing the "matching concordances" activity that was discussed above. Instead of needing a native speaker present, the learner or group of learners could use the corpus as means to validate their answers. Similarly, if a learner is analyzing a word to figure out its semantic range or connotative meaning, he can use a corpus. For example, if the learner is trying to figure out the register of the word *dude*, he can put that into the word search box and have the system display the results as a chart. Results will that the word *dude* only occurs .40 times per million in academic texts, 22 times per million in fictional texts, and 4.5 times per million in spoken language. By looking at this data, the learner should be able to deduct that the word *dude* is not a formal word. Furthermore, by looking at a specific concordance, the learner should be able to make other judgements about the semantic range and connotative meanings of words.

A concordance can also be used by the learner as she is writing. Often times a language learner will use a thesaurus during her writing process and choose words that she feels are more intellectual or advanced. The result is a word that does not fit into its contextual surroundings. If a learner chooses to use a thesaurus, she can use the a concordance to check how the word is used and in what medium it is being used.

The main benefit of the the concordance is that the learner can see how words are being used in real-life language. It can supplement or even supplant the need for a native speaker to help judge the acceptableness of a word or collocation. When one starts using a corpora like COCA, the usefulness surpasses that of a standard concordance. It allows the learner to statistically and visually learn about the use and semanticity of a word. It allows the user to see how words relate to other words of a given word class and regulate the proximity of the words.

Although concordances are great tools, there are drawbacks. One drawback is the accessibility of concordances. This is mainly an issue because they require access to technology, whether they are web-based or actual software that you need to download. Also, the corpora programs that were reviewed in this paper, they were too limited or had such a large number of options that it seemed intimidating. The BNC without a subscription is so limited that it inhibits its usefulness to the learner, and the COCA can do so many things, that it takes time to learn how to do all of them.

5.3.5 Find the Intruder

This activity was developed by Vasiljevic (2008, 46). It involves giving the learner a target word with some acceptable and unacceptable collocates. An example might be: *STRUM*

guitar violin cello banjo harp. The learner then has to decide which of the words are acceptably collocated with the word strum. This type of activity can be used in conjunction with reading that is done in the classroom. If the instructor is using a vocabulary list that coincides with a chapter, this would be a place to begin. Or, if instructor analyzes a piece of writing, this would be another way to narrow the word choices. This activity would also allow

the instructor to assess student understanding of certain lexical items. The use of concordances would also be helpful with this activity.

The benefits and limitations are the same as previously mentioned activities. This is a good way to get a rich understanding of word and how it collocates with other words. It gives the learner multiple exposures to the words. The type of activity would also be good to do in small groups. It has the ability to generate productive and introspective conversation. Since this is directly teaching certain vocabulary words, the number of words that are able to be covered are limited. Although the initial time to complete an activity like this is minimal, the verification of the choices takes considerable time, either through the use of something like a concordance or instructor led discussions.

5.3.6 Focus Paraphrase

Another activity suggested by Vasiljevic (48) is a speaking activity that has students working in pairs. They are given two sets of words. The sets can be constructed of adjective-noun or verb-noun combinations. The learner is given the first sentence as a prompt. The learner then has to decide which words can be successfully collocated in order to make a new utterance that is both acceptable and semantically similar. Both the nouns and the adjectives are synonyms, so it forces the learner to focus more closely on the narrower definitions and connotational colorings of the words. The following examples are given by Vasiljevic (48):

Table 10
Example of a Focus Paraphrase Exercise

Adjectives	Nouns
strong heavy big	vehicles transportation traffic

There were many cars on the street that day-->Traffic was heavy that day.

Another example:

Verb	Noun
hit struck smacked	watch timepiece timer

When I noticed that it was one o'clock, I knew I was late--->When the clock struck one, I knew I was late.

The benefits of this activity are that it helps give the learner a limited number of choices in creating a paraphrase. Unlike activities where the learner is classifying words by their connotative colorings, this activity provides some boundaries for the learner, but at the same time giving the leaner opportunity to learn about nuances of the words. This activity would also work as an introduction to a larger class lesson. After the students complete this activity, they could then break into small groups and discuss the possible answers. As a large group, the students could compile all of the different answers people generated and then discuss if they are acceptable or not. When this is done, there could be a master list that is produced. To take it one step further, if this is an on going exercise, each student could have their own little collocation dictionary where they can record all of the acceptable combinations that have been created by the class. Limitations, once again, are the same as with any activity where learners are focusing on a limited number of vocabulary items.

In summary, these activities were presented with advanced language learners in mind. In this study, it was identified that ALLs had difficulty with collocations and using words that fit both denotatively and connotatively. Two approaches were taken to address these issues. The

first was a primarily implicit and indirect teaching of collocations. This involved extensive reading and listening with a focus on developing an ability to chunk parts of language. Although there was a part of this activity where learners were to mark the natural breaks in the spoken language, it did not focus directly on the the learning of collocations. This type of activity relied on internalization of the language.

The second, and more common, approach to teaching collocations was to explicitly teach certain collocations. The collocations chosen to be taught came from either the student's writing or the instructor. The activities provided the learner with a limited choice of what words to collocate, used a concordance or instructor as a measure of acceptability, or a combination of the two. The major limitation of the second approach is the number of words in the English language and the amount of time it takes to teach few words. However, since these are advanced learners, it would seem appropriate to to say that their breadth of vocabulary is strong, so it is important to take time to develop its depth.

5.4 Conclusion

The word is perhaps one of the most important parts of language. It is the unit of language that most effectively carries the meaning of a message. However, little academic time has been spent analyzing the types of mistakes learner make with their word choice. A better understanding of these mistakes can help guide instruction. This is the ultimate purpose of this research, to focus in on the the mistakes learners are making and provide instruction to meet their specific needs. First, one must understand several aspects of language and instruction. One must understand what a word is and that the answer to this question is not a simple one. There are many views on what constitutes a word, whether it is defined graphically or by units of meaning. Next, it is

important to understand what it means to know a word. There are many levels to this. People can have a word in their receptive vocabulary, which means they recognize it and have an idea of its definition when it is seen or heard. People also have productive vocabulary, which means they are able to correctly use the word in speaking or writing. However, the delineation between these two types of vocabulary is not a clear line, but rather a blurred transitional area. Furthermore, vocabulary moves to and from the productive and receptive vocabulary. Quite often if there is a lexical item that has not been used recently or frequently it will move from the productive vocabulary into the receptive.

In addition to knowing what a word is and what it means to know a word, it is important to have an understanding of how we learn and store language. This paper takes the stance that a language can be acquired at a young age; however, once an individual moves past the critical period, language is learned rather than acquired. It is also the position of this paper that language is largely stored in chunks, beginning at the phonological level and all the way up to the phrasal level. This is believed to be the case because of the studies that have been done with the short term and long term memory. Although the position of this paper is that language is stored in chunks, it does not believe that this inhibits the natural creative ability of language.

Similarly, it is important to understand different methods of vocabulary instruction in order to make informed decisions about how to instruct various learners. This paper focused on the differences and benefits of explicit and implicit instruction. It also looked at the differences and benefits of deductive and inductive instruction. All four approaches have their strengths, but no one method should be exclusively used. When deciding on activities, one should have a balance of these approaches in an attempt to reap each one's benefits. It is also important to give

learners multiple exposure to the word and allow them opportunities to experience the word in all four skills.

Finally, in order to best serve the learner, one must understand the needs of the learner. At each level of language development the learner has general specific needs. The same is true for the advanced language learners. Although advanced language learners have a firm grasp on the mechanics, the grammar, and, in general, the vocabulary of a language, they still have not obtained a native-like status. ALLs no longer need intensive instruction on the grammar points of the language, nor do they need to be inundated with vocabulary. Rather their needs are more specific and have a narrower focus. In grammar, they need to focus on the exceptions, and in vocabulary, they need to focus on the connotative meanings of words. They need to have a richer understanding of the words and how they collocate with other words. These two things, a richer understanding of words and how they collocate, work hand-in-hand and are heavily intertwined. One cannot be accomplished without the other.

In conclusion, this paper has examined the lexical errors in advanced language learners. More specifically it focused on four research questions. First, what types of lexical errors are advanced learners making in their writing? Second, assuming that the study yields findings similar to other identical studies, are the semantic lexical error a result of L1 collocational knowledge? Third, or are the semantic lexical errors the result of a lack of knowledge about content words? Fourthly, what are pedagogical methods to deal with the identified lexical errors, taking into consideration their sources?

Through a lexical analysis of advanced learner compositions it was found that lexical errors made up more than half of all other errors of advanced language learners, which reaffirms

the importance of study in this area. The most common types of lexical errors (37%) had to do with the leaner using words that fit the context denotatively, but not connotatively. The second most common errors the advanced learners made were the totally wrong word, incorrectly paraphrasing, and misallocations. With the use of a survey and native Russian and Korean informants, it was found that there was not a strong correlation between semantic lexical errors and L1 collocations or word coverage. As a result of the semantic lexical errors not being influenced by the L1, it is hypothesized that errors can be attributed to a limited understanding of the full semantic range the words. Finally, it was suggested that vocabulary instruction focuses on the richer understanding of certain lexical items and the teaching of collocations would most benefit advanced language learners.

At the end of the study, certain limitations are recognized. The sample text was fairly small and the demographics of the subjects fairly limited. It would be interesting to expand the corpus used both numerically and demographically and conduct the research again. It would be interesting to see if the study yielded the same results when a variety of subjects were used from varying lingual backgrounds. Another limitation was the analysis of the the essays. In the current study, it was the judgement of acceptability of one native English speaker. Perhaps, as this study is replicated in the future, the addition of one or two other assessors would be beneficial to the validation of this study. Similarly, administering the surveys to multiple informants would also help with the validation of these findings. There was only one informant for the Ukrainian surveys and three for the Korean surveys. It would be helpful to have numerous informants to get a more objective view on the data.

As stated at the beginning of this paper, although there were specific questions it aimed to answer, it was also an exploratory study. At the completion of the study several new topics for further study arose. First, knowing what classifications (verbs, nouns, adjectives, adverbs, etc) learners struggled with most would help guide vocabulary instruction. Secondly, as an extension of this research, it would be interesting to see how effective each of the presented activities are with advanced language learners.   Also, further research could be done on the effect of a language learner's first language on lexical errors and how those errors would be distributed. Finally, further research could be done on the individual constituents of miscollocations (preposition-verb, noun-verb, etc.).

Works Cited

Aebersold, Jo Ann, and Mary Lee Field. *From Reader to Reading Teacher*. Cambridge:

    Cambridge University Press, 1997. Print.

Bahns, Jens and Moira Eldaw. "Should We Teach EFL Students Collocations?" *System* 21.1

    1993: 101-114. Print.

Baumann, James F. and Edward J. Kameenui. "Research on Vocabulary Instruction: Ode to

    Voltaire." *Handbook of Research on Teaching the English Language Arts*. Ed. James

    Flood; Julie M Jensen; Diane Lapp and James r Squire. New York: Macmillan, 604-631.

    Print.

Chang, Yu-Chia. "An Automatic Collocation Writing Assistant for Taiwanese EFL Learners: A

    Case of Corpus-Based NLP Technology." *Computer Assisted Language Learning*. 21.3

    2008: 283-299.

Cowie, A.P. "Phrasal Dictionaries: Some East-West Comparisons." *Phraseology: Theory,*

    Analysis, and Applications. A.P. Cowie. Oxford: Oxford University Press, 1998.

    209-228. Print.

Dekeyser, Robert. "Implicit and Explicit Learning." *The Handbook of Second Language*

    Acquisition. Ed. Catherine J. Doughty and Michael H. Long. Blackwell, 2005.

    313-348. Print.

Dordick, Michael. "Testing for a Hierarchy of the Communicative Interference Value of ESL

    Errors." *Systems* 24.1 1996: 299-308. Print.

Ellis, Nick C. "Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language

    in Good Order." *The Quarterly Journal of Experimental Psychology*, 49A 1996:

    234-250. Print.

Ellis, Nick C. "Phonological Memory, Chunking, and Points of Order." *SSLA*. 18 1996: 91-126.

    Print.

Gathercole, S.E. and A.D. Baddely. *Working Memory and Language*. Hove, England: Lawrence

    Erlbaum Associates Ltd., 1993. Print.

Granger, Sylviane. "Prefabricated Patterns in Advanced EFL Writing: Collocations and

    Formulae." *Phraseology: Theory, Analysis, and Applications*. A.P. Cowie. Oxford:

    Oxford University Press, 1998. 145-160. Print.

Hatch, Evelyn and Cheryl Brown. *Vocabulary, Semantics, and Language Education.* New York:

    Cambridge University Press, 1995. 369-399. Print.

Howarth, Peter. "The Phraseology of Learners' Academic Writing." *Phraseology: Theory,*

    Analysis, and Applications. A.P. Cowie. Oxford: Oxford University Press, 1998.

    161-186. Print.

Hyland, Ken. "Academic Clusters: Text Patterning in Published and Postgraduate Writing."

    International Journal of Applied Linguistics 18.1 2008: 41-62. Print.

Jackson, Howard and Etienne Ze Amvela. *Words, Meaning, and Vocabulary*. London:

    Continuum, 2000. Print.

Kim, YouJin. "The role of Task-Induced Involvement and Learner Proficiency in L2 Vocabulary

    Acquisition." *Language Learning* 58.2 June 2008: 285-325. Print.

Lennon, Paul. "Error: Some problems of Definition, Identification, and Distinction." *Applied Linguistics* 12 1991: 180-195. Print.

Llach, Maria Pilar Agustin. "The Relationship of Lexical Error and their Types to the Quality of ESL Compositions: an Empirical Study." *Porta Linguarum* 2005: 45-57. Print.

Llach, Maria Pilar Agustin. "Lexical Errors As Writing Quality Predictors." *Studia Linguistica* 61 2007: 1-19. Print.

Laufer, Batia. "Development of Passive and Active Vocabulary in a Second Language: Same or Different?" *Applied Linguistics.* 19.2 1998: 255-271. Print.

"Lexicon". *Oxford English Dictionary Online.* Oxford University Press. n.d. Web. 13 August 20012.

Meara, Paul and Fiona English. "Lexical Errors and Learners' Dictionaries." London University 1987: 1-12. Print.

Mel'cuk, Igor. "Collocations and Lexical Functions" *Phraseology: Theory, Analysis, and Applications.* A.P. Cowie. Oxford: Oxford University Press,1998. 23-54.

Mikulecky, Beatrice S. *A Short Course in Teaching Reading Skills*. Reading, MA: Addison-Wesley, 1990. Print.

Moon, Rosamund. "Frequencies and Forms of Phrasal Lexemes in English." *Phraseology: Theory, Analysis, and Applications.* A.P. Cowie. Oxford: Oxford University Press, 1998. 79-100. Print.

Nation, Paul. "Vocabulary Size, Growth, and Use." *The Bilingual Lexicon*. Ed. Robert Schreuder and Bert Weltens. Amsterdam: Benjimins, 1993. 115-135. Print.

Nation, Paul. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University

    Press, 2001. Print.

Nesselhauf, Nadja. "The Use of Collocations by Advanced Learners of English and Some

    Implications for Teaching." *Applied Linguistics* 24.2 2003: 223-242.

O'Rourke, Joesph. *Techniques of Teaching Vocabulary*. Palo Alto: Field Educational

    Publications, 1971. Print.

Robinson, Peter J. "A Rich View of Lexical Competence." *ELT Journal*. 43.4 October 1989:

    274-282. Print.

Salem, Llana. "The Lexico-Grammatical Continuum Viewed through Students Errors." *ELT

    Journal* 61.3 2007: 211-219. Print.

Saville-Troike, Muriel. "What Really Matters in Second Language Learning for Academic

    Achievement?" *TESOL Quarterly* 18.2 1984: 199-219. Print.

Singleton, David. *Language and the Lexicon: An Introduction*. London: Arnold, 2000. Print.

Shahheidaripour, Gholamabbass. "Interlanguage Development and Collocational Clash." *Iran

    Language Institute*. 2000. Print.

Shin, Dongkwang and Paul Nation. "Beyond Single Words: The Most Frequent Collocations in

    Spoken English." *ELT Journal* 62.4 2008: 339-348. Print.

Swain, Merrill and Wataru Suzuki. "Chapter 39. Interaction, Output, and Communicative

    Language Learning." *The Handbook of Educational Linguistics*. Bernard Spolsky and

Francis M. Hult. Hong Kong: Wiley-Blackwell, 2008. 557-580. Online. "factor".

Thesaurus.com. Web. January, 12, 2013.

Vasiljevic, Zorana. "Developing Collocational Competence of Second Language Learners." *The East Asia Learner* 4 May 2008: 46-50. Print.

Wei, Yong. "Teaching Collocations for Productive Vocabulary Development." City University of New York 1999. Print.

Xio, Richard and Tony Mcenery. "Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective." *Applied Linguistics*. 27.1 2006: 103-129.

Zimmerman, Rudiger. "Classification and Distribution of Lexical Errors in the Written Work of German Learners of English." *Papers and Studies in Contrastive Linguistics* 21 1986: 31-40. Print.

Zimmerman, Rudiger. "Form-Orientated and Content-Orientated Lexical Errors in L2 Learners." *International Review of Applied Linguistics* 25 1987: 55-67. Print.