2017

The Frequency of the Twelve Verb Tenses in History Papers Written by University Native Writers

Bushra Qahtani
University of Central Florida

STARS Citation

University of
Central
Florida

STARS
Showcase of Text, Archives, Research & Scholarship

THE FREQUENCY OF THE TWELVE VERB TENSES
IN HISTORY PAPERS WRITTEN BY UNIVERSITY
NATIVE WRITERS

by

BUSHRA QAHTANI
B.A., Taif University, 2014

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Arts
in the Department of Modern Languages and Literatures
in the College of Arts and Humanities
at the University of Central Florida
Orlando, Florida

Fall Term
2017

Major Professor: Keith Folse

ABSTRACT

Due to the variety and complexity of verb tenses in English, English as a Second Language (ESL) / English as a Foreign Language (EFL) learners often experience difficulty in mastering English verb tense system. This corpus-based study was conducted to ascertain the most frequently used verb tenses by English native speakers (NSs), specifically in their academic writing in a history course. As this study aimed to examine the naturally-occurring language produced by NSs, specifically in their academic writing, an original corpus of 101,713 words was assembled. The corpus consisted of 130 research papers written by 65 students. The corpus was analyzed, targeting certain linguistic items: the twelve verb tenses, modals, perfect modals, and imperatives. These targets were highlighted using a code-coloring method. Then, the items were calculated using Microsoft Excel. Excel calculations revealed the number of occurrences of each of the verb tenses, modals, perfect modals, and imperatives that was utilized in the corpus. The results revealed that the simple past tense was predominant, followed by the simple present tenses. Modals came third, as they occurred more frequently than any of the other verb tenses. The remaining targets had percentages ranging between 1.62% and zero. These findings could contribute in developing the methodology of teaching verb tenses to ESL/EFL learners in a way that reinforces their comprehension of the most important and most common items. In addition, this information is very important for materials designers and curriculum professionals.

This work is dedicated with love and gratitude to my father who has been my first and greatest teacher of English, my mom who has been the strongest pillar in my life, my husband who has been my constant source of support and encouragement, and my three-year-old prince, my son Abdullah. It is also dedicated to my uncle, Abdullaziz, my siblings, friends, and parents-in-law, who have motivated and encouraged me throughout the whole journey.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Introduction

When second/foreign language learners learn a language, they must acquire the vocabulary of that language. However, there is no use of vocabulary without grammar. Grammar is what holds words together to make sense of expressions. A grammar point that concerns all English as a Second Language (ESL)/ English as a Foreign Language (EFL) learners throughout their whole journey of learning the English language, from complete beginning levels to the most advanced levels, is learning English verb tenses.

Statement of the Problem

When it comes to teaching English grammar, there appear several teaching points that ESL/EFL teachers need to teach in accordance with the learners' proficiency levels. As teachers follow the assigned curricula and teach the grammar topics that match the learners' levels, it should be taken into consideration that grammar points differ in their importance and relatedness to Non-Native Speakers' (NNSs') language needs, and that these more important grammar points need an extensive effort from both ESL/EFL teachers and learners. English verb tenses, which count twelve, shape an overwhelming grammar point for most ESL/EFL learners, essentially because they are twelve verb tenses. In other words, for many English NNSs, twelve verb tenses in a language is incredibly tremendous. For instance, how tremendous the number of verb tenses in English is in the perspective of an Arabic Native Speaker (NS) learning English as a second/foreign language as Arabic has only two verb tenses!

1

Purpose of the Study

This study focuses on English verb tenses as an important, relevant grammar point to all ESL/EFL learners, aiming to pay a close attention to the importance of each single verb tense to the learners' language needs. Studying the natural language use of English by NSs, the researcher would be able to determine which verb tenses are more important and common in use than the others. Therefore, the researcher used a linguistic corpus of over 100,000 words, collected from NSs' naturally written language, analyzed all the verb tenses used in it, and determined the consequential order of the twelve verb tenses from the most to the least frequent ones. For the sake of strengthening the study goals, the researcher included modals, perfect modals, and imperatives to the targeted linguistic items as they are verb forms, functioning as verbs, but are not considered verb tenses. Investigating the frequency of the twelve verb tenses, modals, perfect modals, and imperatives in the naturally occurring language by NSs could help inform TESOL educators about which of them need more assertion in ESL/EFL curricula.

Research Question

This study is guided by the following research question:

- What are the most frequently used verb tenses that native writers use in assignments in university history courses?

CHAPTER 2: LITERATURE REVIEW

Corpus Linguistic Studies

Corpus-based studies have been widely used in the field of Teaching English for Speakers of Other Languages (TESOL). Corpus linguistics research is a "methodology for linguistic analysis that focuses on describing linguistic variation in large collections of authentic texts (the corpus), using automatic and interactive computer programs to aid in analysis" (Gray & Biber, 2011, p. 139). The goal of corpus-based studies is to illustrate the lexical and grammatical use of language in various situations. This kind of research uses computers to analyze the collected corpus while previous studies relied on human intuition or manual calculations. Therefore, the findings of corpus-based studies are quite reliable and are comparable.

Corpus-based research has made great contributions in building up TESOL educators' knowledge of the variation in lexical and grammatical uses of English. The following are some examples of corpus-based studies that are great resources to inform ESL teaching.

Examples of Some Corpus-Based studies:

Language Lexical Features

Biber, Conrad, and Reppen (1998)

Generally, most corpus-based research revolves around studying the language in depth in terms of the lexical and/or grammatical uses of the language. In 1998, Biber, Conrad, and Reppen published their book, *Corpus linguistics: Investigating language structure and use*, in

which they focused on the language use in both speech and writing. Using a corpus-based approach, the book discussed in each chapter a linguistic area studied by analyzing a corpus of large databases. The book proved the usefulness of corpus-based studies in terms of understanding the language use.

Lexicography, for example, is the first chapter of the book in which the authors discussed studying meaning, use, and synonyms of words for the purpose of making dictionaries. This chapter provided some corpus-based studies that sought some research questions about the meanings associated with particular words, the frequency of a word compared to that of other related words, the non-linguistic associations (e.g., registers) that a ward has, a word's collocations and their distribution across registers, the senses and uses of words, and the different ways synonyms are used and distributed. (Biber et al., 1998, p. 23-24).

The authors used these research questions as a guide of the chapter. They answered each question based on a corpus-based study. For example, the last question was to investigate how synonymous words are used in variant ways. The authors considered three synonyms of size (*big, large, and great*) by investigating their frequency distribution, immediate right collocates, and collocates of at a larger distance.

Using a robust sample of 5.7 million words from the Longman-Lancaster Corpus, the researchers found extreme variation in the frequency distribution of these three words between two registers: academic prose and fiction. In fiction, *big* and *great* were notably used much more often than *large*. However, in academic prose, the word *large* was three times more frequent than *big* and *great*.

In addition to the relative frequency of these three words, another important question for learners, teachers, test developers, and materials writers is the words that appear to the immediate right of each of these adjectives, i.e., which nouns are most likely to combine with each of these apparently synonymous adjectives. To determine these collocations, Biber et al. (1998) used samples of academic prose and fiction from the Longman-Lancaster Corpus as well. The authors created a table of the most common right collocates of these three words. The authors gave a detailed analysis of the most common right collocates of these words in each register, showing that synonymous words are not equivalent in meaning at all, and that "Each adjective has its own preferred collocates, different preferred sense, and different distributions across registers" (Biber et al., 1998, p. 51). The following table (Table 1) shows some of the frequent right collocations of the words (*big, large,* and *great*) that the authors found in the two registers. The table presents the top most common right collocations of each word.

Table 1:Examples of the Most Frequent Right Collocations of (Big, Large, and Great) from the Study of Biber et al. (1998)

	BIG	LARGE	GREAT
Academic prose	Enough	Number/ numbers	Deal
	Traders	Scale	Importance
		And	Number
		Enough	Majority
Fiction	Man	And	Deal
	Enough	Black	Man
	And	Enough	Burrow
	Black	House	Big

Source: (Biber et al., 1998, p.46).

The last part of the authors' answer for that last question was to investigate the collocates that are of a large distance. To do so, Biber et al. (1998) aimed to find the second word to the right of the word *large*. It was found that the collocational *large X of* is the most common one in

both registers, meaning that *of* is the most common collocate at a distance. The authors also found that *large* often takes the frame *large+ N+ of* (N= noun of amount).

It is beyond the scope of this literature review to present all of the questions presented in this first chapter, nor to indicate all examples of corpus-based approach that the authors used. However, this information for *big*, *large*, and *great* is given to show how Biber et al. (1998) demonstrated the power of corpus-based analysis in terms of helping language teachers and curriculum designers make vocabulary teaching representative of its natural occurrence. In sum, language professionals no longer had to rely solely on their intuition and guesswork and instead had actual real-world data presented in this work by Biber et al. (1998).

Coxhead (1998, 2000)

One of the most welcomed applications of corpus linguistics by teachers and learners alike is the creation of word lists. Today, there are many word lists of vocabulary resulting from corpus linguistics, as Lessard-Clouston's (2013) and Youngblood and Folse (2017) have documented. Perhaps the most well-known of these lists is Coxhead's (1998, 2000) pioneering Academic Word List (AWL), which is a list of 570 academic word families that represented the most commonly used words in various academic texts. Coxhead (1998) used a 3.5 million-word corpus of collected academic written texts from four academic sections: Arts, Commerce, Law, and Science. Coxhead (1998, 2000) analyzed the corpus and looked for the words that occurred over 100 times in the overall corpus, that occurred 10 times at least in each academic faculty section, and that were not included in the General Service List (GSL).

The AWL has become very popular within TESOL. In fact, it is perhaps the leading list for ELLs (English Language Learners) studying English for Academic Purposes (EAP). It seeks to simplify the task of vocabulary learning for ELLs since it provides them with the words that are truly necessary to know because these words are highly likely to be encountered in an academic context.

In her research, Coxhead (1998, 2000) used a 3.5-million-word corpus to lead the AWL to meet the principles of corpus linguistics in terms of creating word lists. In addition, she examined the coverage of the AWL of the overall corpus and of its four sections. Coxhead (2000) evaluated the AWL by examining its coverage of the corpus along with the GSL, using a second corpus (a second collection of academic texts), and using a corpus of fiction texts. The AWL is a specialized vocabulary list that provides good coverages within a corpus of academic texts, regardless of the subject area. Therefore, this list can be used in EAP courses as a great resource for teaching and learning. (Coxhead, 2000, p. 226-227).

Chen and Ge (2007)

Believing in the importance of Coxhead's AWL, Chen and Ge (2007) studied the text coverage and the word frequency of the 570 words Coxhead's AWL in 50 medical research articles (RAs). The corpus used in this study was obtained from the public Internet and two databases: ScienceDirect Online and Medline. The researchers found that 10.073% was the text coverage of AWL in the collected RAs, which corresponds with what Coxhead and Nation (2001) reported. In addition, Chen and Ge (2007) found out the most frequently used academic words form the AWL in the RAs, considering the words that occurred more than 10 times as

frequently used academic words. The researchers found that only 292 word families, among the 570 AWL word families, were frequently used in the collected corpus. The results showed that *significance, analysis, data,* and *factor* are very frequently used; however, *ratio* was the most frequently used AWL word among all.

An important conclusion of this study was Chen and Gu's (2007) discussion of the differences between the frequency of AWL words in Coxhead's corpus and in their corpus. It was apparent that not all AWL words are frequently used specifically in the field of the corpus used in this study. Some AWL words that were viewed as most frequently used words (e.g., *formula, legal, policy,* and *economy*) were not as frequent in Chen and Ge's (2007) study. On the other hand, some unfrequently used AWL words in Coxhead's corpus (e.g., *media, logic, found, mode, detect,* and *induce*) were found more frequently used in the corpus of this study. Also, Chen and Ge (2007) found some AWL words that did not occur at all in their corpus including *community, authority, labor* and *financial*.

Vongpumivitch, Huang, and Chang (2009)

In their study, Vongpumivitch, Huang, and Chang (2009) sought the frequency of AWL words utilized in applied linguistics research articles. They also aimed to identify content, non-AWL words that were found used with high frequency in these articles. For this study, they created a large corpus of 1.5 million words called the Applied Linguistics Research Articles Corpus (ALC). The corpus was a collection of 200 journal articles published in five different applied linguistics journals, 40 research articles from each.

To explore the frequency of AWL words, the researchers set two criteria to create a list of the AWL words used frequently in the corpus. There was the frequency criterion (i.e., an AWL word should occur at least 50 times) and the range criterion (i.e., an AWL word should occur at least five times in each journal). Based on these criteria and using computations, Vongpumivitch et al. (2009) were able to calculate the frequency of AWL words utilized in ALC, and then provide a ranking of them.

Furthermore, Vongpumivitch et al. (2009) used the same criteria (i.e., frequency and range) to create a list of content, non-AWL words. Their goal was to explore the content words that are used frequently in the field of applied linguistics, but are not included in the AWL list. The researchers used computations to count these words, excluding function words such as conjunctions, prepositions, etc.

The results revealed that AWL words were used in the corpus with a coverage of 11.17% of the whole ALC, which verified the great role AWL words play in the field of applied linguistics. Specifically, the researchers found 475 AWL words meeting the criteria of frequency and range that the authors set, with a percentage of 8.6%. In other words, these 475 words occurred at least 50 times, and at least five times in each of the five applied linguistic journals). The results, furthermore, identified 128 content, non-AWL words the met the authors' criteria. These non-AWL words made a percentage of only 2.8%, which is, interestingly, much less than the coverage of the AWL words. In sum, although it is useful that learners know the non-AWL words that are frequently used in applied linguistics, words in the AWL list are still with a great importance. (Vongpumivitch et al., 2009, p. 36-38).

9

<u>Liu (2012)</u>

Liu (2012) also examined academic English, specifically for the purpose of identifying the most frequently-used multi-word constructions (MWCs) in academic writing. The MWCs considered here included lexical bundles (LBs), idioms, and phrasal/prepositional verbs. In this study, the academic writing sub-corpora of the Corpus of Contemporary American English (COCA), and the British National Corpus (BNC) were used. The data used in this study was collected from 150 different academic journals across different academic divisions. Liu (2012) aimed to examine the most common MWCs, the functions they play, and the differences in their use between American and British academic English. In addition, he searched the extent to which the findings of previous studies about the use of MWCs can be certified in large corpora.

The study resulted in a list of 228 most frequently-used MWCs in in academic written English. The items in the list were grouped according to their frequency and semantic function. Regarding the previously conducted research, the results of this study confess earlier findings. For example, the results of this study confirm what other studies found that; for example, noun and prepositional constructions are the two largest types of MWCs, and that MWCs are discipline-specific. In addition, the results provided new findings. For instance, some verb phrases such as *tend to* and *take place* are very common in academic writing. Based on these findings, Liu (2012) also examined differences between American and British academic written English. Liu (2012) found that some MWCs were mostly found in American academic writing while others are mostly found in British academic writing. For example, *as long as* and *keep in mind* are found mostly in American English. on the other hand, *as far as* and *bear in mind* are found mostly in British English.

<u>Liu (2003)</u>

In another research study, Liu (2003) claimed that teachers and materials often refer to rarely used idioms or describe their meanings incorrectly (p. 671). To examine this issue empirically, the researcher used a corpus-based study to demonstrate the most frequently spoken American English idioms used by English for Speakers of Other Languages (ESOL) students learning American English. The study searched and analyzed the idioms used in three large spoken American English corpora: Corpus of Spoken Professional American English (Barlow, 2000); Michigan Corpus of Academic Spoken English (Simpson, Briggs, Ovens, Swales, 2002), and Spoken American Media English (Liu, 2002). The study resulted in four lists of the most commonly used idioms: one for each corpus and one for the combined corpora. The study also proved that pure idioms are rare, register sensitive, variant in form and tense.

Liu (2003) compared the results of this study with seven dictionaries and two idiom publications (nine materials) and found that the selection of idioms in these materials is inconsistence. For example, the dictionaries include idioms such as *above par, all over again, and all right*, but not *according to* which the researcher found to be much more frequent than *above par*. Another highly frequent idiom, based on the results of this study, is *with/in regard/respect to* which is found to be mentioned in only one dictionary.

Moreover, Liu (2003) compared the meanings and usage explanations of the idioms presented in the materials with the meanings and usages of these idioms in the corpora. The researcher gave an example of the verbal idiom *to bring up*. The materials that include this idiom gave a primary meaning that is *to rear or educate a person (often a child)*, and a secondary or third meaning that is *to mention or start discussing an issue*. However, by analyzing the

collected corpora, it was found that the second meaning makes up more than 90% of the idiom uses whereas the primary meaning constitutes only 5%.

Another point that Liu (2003) referred to is that the materials do not introduce the variations in many idioms, nor do they introduce their frequency. For example, some dictionaries present *with regard to* and *in regard to* without referring to their frequency which, in the corpora, is found to be vastly different (the former is seven times more frequent than the latter). However, *with respect to* is the most highly frequent phrase when compared to the other variations with/in regard to and in respect to.

This study and others like it can inform ESL teachers about the most frequent idioms used by American English speakers and their most common uses. ESL teachers and ESOL students can use corpus-based lists also to indicate the idiom variations.

Simpson and Mendis (2003)

Another corpus-based research that studied the usage of idioms, particularly in Academic speech, is Simpson and Mendis's (2003). In this study, the researchers used a corpus of 1.7 million words of academic discourse, using the Michigan Corpus of Academic Spoken English (MICASE). It included 197 hours of recorded speech (e.g., lectures, office-hours conversations). The corpus was divided based on the primary discourse mode into three categories: monologic/ panel, interactive, and mixed, and divided based on the academic division into Humanities and Arts, Social Sciences and Education, Biological and Health Sciences, and Physical Sciences and Engineering. The researchers sought whether or not idioms are used at all in spoken academic English, and, if so, what functions they perform.

Simpson and Mendis (2003) started by creating lists of idioms used in three university-level ESL textbooks designed to teach idioms (Madden and Rohlck, 1997; McCarthy & O'Dell, 1997; and Redman & Shaw, 1999) that were published around the same time the MICASE was assembled. Comparing these lists with their corpus, the researchers found that merely 25% of the idioms in the list are used in the MICASE. This was attributed to the MICASE size and to the unprincipled selection criteria that the textbooks authors used to select and include idioms in their materials. As a result, Simpson and Mendis (2003) returned to define idioms and search the corpus.

The researchers defined an idiom as "A group of words that occur in a more or less fixed phrase and overall meaning cannot be predicted by analyzing the meanings of its constituent parts" (Simpson & Mendis, 2003, p. 423). This definition was the authors' baseline throughout the process of analyzing the corpus. They searched the corpus and used a concordance program to count the relative frequency rate of each idiom.

The findings of Simpson and Mendis's (2003) research study revealed that the use of idioms in spoken academic language is not a rare phenomenon. The researchers claimed that there is a significant number of idioms in the MICASE that they should be included in an EAP curriculum (Simpson & Mendis, 2003, p. 432). In addition, the results revealed no striking differences in the use of idioms neither between the discourse modes (i.e., panel, interactive, or mixed) nor between academic divisions (i.e., University of Michigan graduate school's division of departments).

The findings also showed the various functions that idioms perform in the spoken academic discourse. The following table (Table 2) depicts these functions and provided examples as presented in Simpson and Mendis's study.

Table 2: Functions of Idioms as Presented in Simpson and Mendis's (2003) study

FUNCTIONS OF IDIOMS	EXAMPLES
Evaluation	*Out of whack, throw someone for a loop.*
Description	*hand-in-hand, run-of-the-mill.*
Paraphrase	*Put up a stink, no mean feat.*
Emphasis	*Throwing everything in but the kitchen sink.*
Collaboration	*Put heat on somebody.*
Metalanguage	*Cut to the chase, lost someone's train of thought*

Source: (Simpson & Mendis, 2003, p. 427-432)

Simpson and Mendis (2003), furthermore, presented a list of 32 idioms that were found frequently used in MICASE, including *bottom line, the big picture, come into play, what the hell, down the line, what the heck*. They also presented a list of idioms that they thought would be particularly useful for EAP curricula, including *bottom line, the big picture, carrot and stick, chicken-and-egg question, come into play, draw a line between*, etc.

This study proved the power of corpus-based studies. In this study, "MICASE provides a rich resource for teaching materials that allow teachers not only to use authentic, attested examples of idioms in context but also to consider larger issues of discourse and sociopragmatics" (Simpson & Mendi, 2003, p. 437).

Kennedy (2003)

Using the British National Corpus (BNC), Kennedy (2003) studied adverbs of degree (amplifiers) in terms of the way they collocate with particular words. The study focused on the

24 most frequent amplifiers used in the BNC, subcategorized in two categories: maximizers and boosters. In this study, Kennedy (2003) identified eight maximizers (*fully, completely, entirely, absolutely, totally, perfectly, utterly,* and *dead*) and 16 boosters (*very, really, particularly, clearly, highly, very much, extremely, badly, heavily, deeply, greatly, considerably, severely, terribly, enormously,* and *incredibly*). After identifying these as the most frequently used lexical items, Kennedy (2003) then retrieved the collocations for these 24 amplifiers and the words they modify, that is, two words on each side of the amplifier.

The results of the study demonstrated that amplifiers are not interchangeable. Some amplifiers do not sound acceptable/comfortable when used with particular adjectives, so they are not likely to occur in the corpus. The results also provide the 40 most frequent collocations associated with the selected maximizers. Apparently, each maximizer collocates with different words. Adjectives and verbs, on the other hand, tend to be associated with particular maximizer. For example, *fully* is exclusively associated with positive collocations (e.g., *fully fledged,* and *fully conversant*) whereas *totally* is mainly collocates with negative collocations (e.g., *totally unsuited,* and *totally unprepared*). In addition, the results provide an enormous number of collocations that are found strongly associated with the selected boosters. Like the maximizers, each booster tends to collocate with different words. For example, *badly* is found particularly associated with damage (e.g., *badly mauled,* and *badly sprained*), *clearly* is associated with perception (e.g., *clearly visible,* and *clearly audible*), *extremely* is associated more with negative adjectives than positive ones (e.g., *extremely difficult,* and *extremely risky*), and *greatly* is associated mainly with positive associations despite some exceptions (e.g., *greatly appreciated,* and *greatly admired*).

<u>Laufer and Waldman (2011)</u>

This study examined the use of noun-verb collocations in the written language, comparing native speakers and non-native speakers as well as comparing non-native speakers with different language proficiency levels. In this study, Laufer and Waldman (2011) used two corpora. One was a corpus of 300,000 words that consisted of argumentative and descriptive essays written by L2 learners. These essays were taken from the Israeli Learner Corpus of Written English (ILCoWE). The other corpus was a corpus of young adult native speakers of English: Louvain Corpus of Native English Essays (LOCNESS).

Laufer and Waldman (2011) started with the LOCNESS scanning all the nouns and making a frequency list for them. After that, they selected the 220 most frequent nouns and created concordances for each noun. Therefore, non-verb collocations can be identified. The last step was checking the extracted non-verb collocations in two dictionaries (*The BBI Dictionary of English Word Combinations* and *The LTP Dictionary of Selected Collocations*) to verify that noun-verb collocations listed in this study were identified as collocations in either one of the two dictionaries.

Laufer and Waldman (2011) also analyzed the learner corpus, extracting the 220 most frequent nouns found in the LOCNESS, making concordances for each noun, and verifying them in the collocation dictionaries. In addition, they analyzed the subcorpora that belonged to the different proficiency levels.

The results of this study indicated that learners, at all their different proficiency levels, underuse noun-verb collocations producing considerably fewer collocations than native speakers.

However, as the three-separate learner subcorpora compared, the number of noun-verb collocations increased significantly at the advanced level although the interlingual errors persisted. Laufer and Waldman (2011) proved the slow development of collocation use, which indicates that using collocations in the written discourse is problematic for L2 learners.

Nesselhauf (2003)

This is another study on collocations in which Nesselhauf (2003) explored the difficulties of using verb-object-noun collocations (e.g., *take a picture* and *shake one's head*) by German-speaking learners at an advanced level. Nesselhauf (2003) distinguished between three classes of word collocations: free collocations (F) when the verb and the noun are both used in unrestricted senses (e.g., *want a car*), restricted collocations (RC) when the noun is used in an unrestricted sense while the verb is used in a restricted sense, i.e., the verb is restricted to particular nouns (e.g., *take a picture*), and idioms (I) when both of the noun and the verb are used in a restricted sense (e.g., *sweeten the pill*). (Nesselhauf, 2003, p. 226). In addition, the researcher used the abbreviation (RC?) to indicate an unclear sense of the verb, i.e., whether it is restricted or unrestricted.

In this study, the researcher used 32 randomly selected essays from the German subcorpus of The International Corpus of Learner English (ICLE). The essays were written by German-speaking L2 learners who were seeking high competency level in English. To analyze the collected data, Nesselhauf (2003) extracted all verb-object-noun combinations from all the essays and classified them into F, RC, RC?, or I using two dictionaries: *Oxford Advanced Learners' Dictionary (OALD)* and *Collins COBUILD English Dictionary (CCED)*. After that,

the researcher examined the acceptability of the combinations produced by the learners in English, considering them correct (C) if they were found in the *OALD,* the *CCED, The BBI Dictionary of English Word Combinations,* the *Oxford Dictionary of Current Idiomatic English,* or found in at least five texts in the *British National Corpus*. The combinations that were not found correct were presented to two native speakers (one American and one British) to judge them as correct (C), wrong (W), or not sure (CW), and then give corrections for the W and CW combinations. Two additional native speakers were presented with the combinations when combinations were judged C and W. This resulted in a five-stage scale of acceptability: clearly acceptable C, largely acceptable {C}, unclear CW, largely unacceptable {W}, and clearly unacceptable W. (Nesselhauf, 2003, p. 230).

From the 1072 verb-object-noun combinations that Nesselhauf (2003) extracted, only 213 were classified as collocations (RC or RC?). In these collocations, nine different types of mistakes were found committed by the learners. However, the most frequently occurring one was the wrong choice of verb (e.g., **carry out races, hold races*). The other types of mistakes included the wrong choice of noun (e.g., **close lacks, close gaps*), using a combination incorrectly (e.g., *take notice, to notice*), using no combination and the combination cannot be corrected by exchanging elements (e.g., **hold children within bounds, show children where the boundaries lie*), using an unacceptable/wrong preposition of prepositional verb (or missing it) (e.g., **fail in one's exam, fail one's exam*), using an unacceptable/wrong preposition of a noun (or missing it) (e.g., **raise the question about, raise the question of*), using an unacceptable/wrong article or pronoun (or missing it) (e.g., **get the permission, get permission*),

18

using a singular noun instead of plural and vice versa (e.g., *pass one's judgments, pass judgment*), and using wrong syntactic structure (e.g., *make sb. friends, make friends with sb.*).

This study also investigated the role of the degree of restriction of a combination. Therefore, Nesselhauf (2003) distinguished between the three major classes of word combinations: F, RC, RC? and I. He; in addition, subdivided the category RC into RC2 (a little restriction) and RC1 (a lot of restriction). The results indicated that the number of mistakes was the highest with (RC2) while it was the lowest with RC1. This finding indicated that L2 learners often acquire and produce collocations where not many nouns can possibly collocate the verb (i.e., RC1 such as *pay attention,* and *run a risk*) than collocations where numerous nouns can collocate the verb (i.e., RC2 such as *exert influence/ control/ pressure/authority/ or power*).

Another finding of this study was related to the role of the learners' L1 on the production of collocations. Nesselhauf (2003) found a considerable influence of the learners' L1 on their mistakes when producing collocations. For example, there is an assumption of L1 influence when a German-speaking learner produces *make homework* because German has *Hausaufgaben machen* where *machen* is related, in meaning and form, to *make*. (Nesselhauf, 2003, p.234). In addition, the researcher found five different types of mistakes in collocations resulted from the influence of L1. These kinds of mistakes were found related to verbs, nouns, usages, prepositions, determiners, number and structure.

Hyland and Milton (1997*)*

The researchers in this study attempted to compare the expressions of doubt and certainty used in academic essays by Cantonese students and NSs to investigate whether NNSs and NSs

writers employ different strategies to modify assertion. They aimed to determine the forms and meanings NNSs and NSs use to express claims in their academic writing.

The data used in this study consisted of two corpora. One was a collection of essays written by Chinese NNSs for the matriculation General Certificate of Education (GCE) A level "Use of English" examination. This corpus consisted of 500,000 words from 150 exam scripts in which the grades ranged from A to F. The other corpus was a collection of 770 GCE A level General Studies scripts written by NSs of the same age and educational level as the Chinese students. This other corpus consisted also of 500,000 words.

To make a comparison between the expressions of doubt and certainty used in both corpora, Hyland and Milton (1997), based on different sources, prepared a list of such expressions common to NSs usage. The list consisted of 75 items of the most frequently used epistemic lexical items in NSs' academic writing. Then, the researchers examined the frequency of these expressions in the two corpora, selecting and extracting fifty sentences that contain each of these expressions from NNSs' corpus (i.e., from each grade) and NSs' corpus sample.

The results of the study indicated that there are remarkable similarities in the overall frequency and usage of the epistemic items in both corpora. Both NSs and NNSs employ one expression every 55 words. Also, both NSs and NNSs use *will, may, would,* and *always* among the six most frequently used expressions. However, an important finding stood out to the researchers was that NNSs used a more restricted range of assertion modifiers. That is, in the NNSs' corpus, the ten most frequently used devices occurred with a total of 75%. This restricted

range of items offer stronger commitments and cause NNSs problems in conveying doubt and certainty in their writing.

The study findings also presented the grammatical distributions of the assertion devices: modal verbs, adverbials, lexical verbs, adjectives, and nouns. Hyland and Milton (1997) found that both NSs and NNSs significantly use modal verbs, particularly *will, would,* and *may*, to modify assertion. However, NNSs depend much more heavily on modal verbs in their essays. On the other hand, NSs have a great frequency of adverbials in their writing. Also, both student groups tend to use significantly much less lexical verbs, adjectives and nouns than adverbials to express doubt and certainty. Comparing students' proficiency variation to grammatical classifications, the findings revealed that advanced NNSs use significantly more devices than lower NNSs.

Furthermore, the study findings revealed that there exist substantial differences between the two student groups in terms of the degree of certainty and tentativeness used. Comparing NSs and NNSs in using epistemic devices to express certainty, probability, possibility, usuality, and approximation, NNSs tend to use 60% more certainty markers (e.g., *always, never*) than their counterpart NSs do. NSs; on the other hand, use 73% more probability markers (e.g., *likely, may*) in their essays. Devices expressing possibility, usuality, and approximation have almost the same distributions.

Language Grammatical Features

<u>Conrad (1999)</u>

To study the grammatical uses of the language, a number of corpus-based studies have

been conducted. One of them is Concrad's (1999) study in which she emphasized the usefulness

of using language corpora to analyze and study the language in depth. In her corpus-based study,

Conrad (1999) studied and analyzed linking adverbials using a 40 million-word Longman

Grammar of Spoken and Written English Corpus, focusing on four different registers:

conversation, fiction, newspaper reportage, and academic prose. Conrad (1999) examined linking

adverbials in terms of frequency, semantic category, grammatical structure, placement within the

clause, the specific item used, and variation across the four registers.

Conrad's (1999) study found that linking adverbials are most common in academic prose,

conversation, fiction, newspaper reportage in order, meaning that academic prose and

conversation are the two registers that have the highest frequencies of linking adverbials.

However, the study also showed how these two registers are distinct in terms of the use of

different semantic categories. Based on the results of Conrad's (1999) study, putting the linking

adverbials used in academic prose in order would be as follows:

1. Result/inference adverbials (e.g., *thus,* and *as a* result).

2. Appositional adverbials (e.g., *for example,* and *that is to say*)

3. Contrast/concession adverbials (e.g., *however*).

4. Enumerative, additive, and summative adverbials (e.g., *in addition,* and *in sum*).

In the conversation register; on the other hand, the result/inference adverbials made the largest proportion of linking adverbials while contrast/concession adverbials came second. *So* and *then* are the most frequently used result/inference adverbials while *though* and *anyway* are the most frequently used contrast/concession adverbials.

The results also shed light on the grammatical structure of linking adverbials used in academic prose versus conversation. Although single-word adverbials are common in both registers, prepositional phrases are common relatively in academic prose. Linking adverbials in academic prose are divers in their grammatical structures while they are repeatedly used in conversation, especially *so, then, anyway, and though.*

In the academic prose and conversation registers, the position of linking adverbials varies. The initial position is the highest percentage in both registers. However, in academic prose, the medial position is common, but the final position is rare. On the other hand, in conversation, the medial position is incredibly rare while the final position is relatively common. This finding alongside with all the findings presented earlier are resulted from using the corpus-based approach which Conrad (1999) believed to be useful to study even complicated grammatical features such as linking adverbials.

Gardener and Davies (2007*)*

In another study, Gardener and Davies (2007) used the BNC aiming to investigate the most frequently used phrasal verbs, attempting to narrow the scope of phrasal verbs in ESL/EFL courses. In addition to this one goal, Gardener and Davies (2007) aimed to:

- Determine the most common adverbial particles (AVPs)

- Determine how common these forms function as AVPs or as prepositions.

- Determine the most common lexical verbs (LVs) used in phrasal verb constructions.

- Determine how often LVs function as stand-alone verbs versus phrasal verbs.

- Determine how common LVs interact with various AVPs.

- Create a list of the most common phrasal verbs (verb-plus-particle) based on frequency and coverage.

- Determine the number of word senses associated with each of most common the phrasal verbs.

After analyzing the corpus, Gardener and Davies (2007) created a table showing the frequency of the AVPs, indicating the number of times each of these forms (e.g., out, up, down, back, etc.) functions as AVP versus as a preposition or another grammatical structure. The results indicated that these forms function as AVPs 15.6% of the time, and that some forms such as *out, up, down*, and *back* act more likely as AVPs in phrasal verbs rather than acting as prepositions. On the other hand, *under, by,* and *across* act much more often as prepositions than as AVPs.

Regarding the frequency of LVs functioning in phrasal verb construction, Gardener and Davies (2007) found that about 5% (i.e., 518,923 LV tokens in PVs) of all LVs in the BNC (10,404,107 LV tokens) function in PV constructions. That is, one in every 20 LVs functions as a part of a PV construction. Pedagogically speaking, ELLs are expected to encounter one PV in every 192 words, or two in every page of written text. (Gardener & Davies, 2007, p. 347).

The study also provided a list of the 20 most frequent LVs involved in PV constructions in the BNC. The list included *go, come, take, get, set, carry, turn, bring, look, put, pick, make, point, sit, find, give, work, break, hold,* and *move*. These verbs were found in more than half (53.7%) of all PVs in the BNC. Additionally, for these 20 most frequent LVs, the study provided the verb-plus-particle constructions, investigating that these 20 LVs combine with eight particles only: *out, up, on, back, down, in, over,* and *off)*. That is, more than half of the PVs in the BNC were basically a total of 160 combinations of these 20 LVs and the eight particles.

The findings also provided a list of the 100 most frequent PVs (verb-plus-particle constructions), besides presenting their frequency and coverage. The list included, for example, *go on, carry out, set up, pick up, go back, come back, go out, point out, find out,* and *come up*. Moreover, the study addressed the semantic characteristics of PVs by indicating the word-sense frequencies for the 100 most frequent PVs in the corpus. For example, *go on* has five meanings, *set up* has 15 meanings, and *break up* has 19 meanings (the highest number of meanings in the list). The findings of this study can be very useful for ESL/EFL teachers in terms of teaching PVs. "[The] high-frequency lists of PVs will partially answer the where-do-we-start question so often asked by English language learners, teachers, curriculum designers, and materials developers" (Gardener & Davies, 2007, p. 353).

Frazier (2003)

Frazier's (2003) study is one that focuses on studying a specific grammatical structure: would-clauses representing a hypothetical and counterfactual meaning without adjacent to if-clauses, a grammatical structure that is rarely indicated in ESL/EFL textbooks (Frazier, 2003).

Frazier's primary aim of this study was to empirically demonstrate the occurrences of

hypothetical or counterfactual would-clauses with adjacent if-clauses and without adjacent if-

clauses. The study also sought to investigate the functions of would-clauses occurring without

adjacent if-clauses and investigate their frequencies. Frazier (2003) used three different corpora:

Brown corpus, Santa Barbara Corpus of Spoken American English (SBC), and Michigan Corpus

of Academic Spoken English (MICASE). The findings of this study indicated that hypothetical

and counterfactual would-clauses occur much more often without being accompanied by if-

clauses, with a percentage of 80%. Would-clauses without adjacent to overt if-clauses have the

following three main functions: nonadjacent if-clauses: conditional frames, no overt conditional.,

and alternative conditionals. Like other corpus-based studies, Frazier's (2003) study contributed

to inform ESL/EFL teachers and materials designers about the authentic use of the language. The

study indicated the authentic uses of would-clauses that seemed to not to be converted as should

be in ESL/EFL textbooks.

Hinkel (2004)

In Hinkel's (2004) study, some verb phrase features (tense [present, past, and future], aspect

[progressive, and perfect], and passive voice) were studied and analyzed, comparing native

speakers' (NS) and non-native speakers' (NNS) usage of them in their academic writing. The

main goal of the study was to determine whether NSs and NNSs write similarly in terms of using

these verb phrase features. 746 students participated in this study (631 NNS and 115 NS). The

corpus used in this study was a collection of 746 essays (226,054 words) written by the students

who speak different first languages (English, Chinese, Japanese, Korean, Indonesian,

26

Vietnamese, and Arabic). Hinkel (2004) had the students write in response to one of the following three prompts:

1. Many people believe that grades do not encourage learning. Do you agree or disagree with this opinion? Be sure to explain your answer using specific reasons and examples.

2. Some people learn best when a classroom lesson is presented in a serious, formal manner. Others prefer a lesson that is enjoyable and entertaining. Explain your views of this issue. Use detailed reasons and examples.

3. Some people use their major field of study based on their personal interests and are less concerned about future employment possibilities. Others choose majors in fields with a large number of jobs and options for employment. What position do you support? Use detailed reasons and examples. (p. 12).

The results of this study revealed that even advanced NNSs may have difficulty using tense, aspect, and passive voice in their academic writing. NNSs employ past tense much more frequently than NS. In addition, NNS's usage of present tense does not make their writing appear academic enough. NNS's written production also indicates the students' lack of awareness of the functions of the future tense in the academic writing (e.g., avoiding using would). Furthermore, the results of this study indicated that most of NNS did not employ progressive and perfect aspects in their writing and that they tended to avoid using passive constructions. That is, NNS tended to avoid any complex verb constructions.

<u>Altenberg and Granger (2001)</u>

This study sought to investigate EFL learners use of the high frequently used verbs including *have, go, take, do, say, look, know, see, give, think, come, find, get, make* and *use*, focusing on the verb *make* and considering it as a major representative of high-frequency verbs. Altenberg and Granger (2001) investigated whether EFL learners over- or underuse high-frequency verbs, how EFL learners' use of these verbs differ from native speakers' use in terms of the verbs' meanings/uses, whether or not the use of these verbs causes EFL learners' errors, and what role does transfer play in the learners' misuse of these high-frequency verbs.

For the purpose of the study, Altenberg and Granger (2001) used an authentic learner corpus that consisted of two corpus samples taken from the *International Corpus of Learner English (ICLE)*: one contained essays written by French-speaking EFL learners (abbreviated as FR) and the other one contained essays written by Swedish-speaking EFL learners (abbreviated as SW). Altenberg and Granger (2001) also used a native speaker corpus as a control corpus: a corpus sample from the *Louvain Corpus of Native English Essays(LOCNESS)* that contained essays written by native speakers.

To figure out whether EFL learners over- or underuse the verb *make*, Altenberg and Granger (2001) extracted all the inflectional forms of the verb *make* (*make, makes, making,* and *made)* from the three corpora (FR, SW, and LOCNESS) using *WordSmith Tools'* lemmatizing facility. The researchers found that the verb *make* is used less frequently in FR (234.6) than in SW (354.3). However, it is used in SW more frequently than in LOCNESS (339.8). That is, compared to the native speakers use of the verb *make*, French-speaking learners underuse it while Swedish-speaking learners overuse it.

The second research question in this study was about the categories of meaning/use that differentiate EFL learners from native speakers. To answer this question, Altenberg and Granger (2001) listed eight major categories of use of the verb *make*: produce something (e.g., *make a hole*), delexical uses (e.g., *make a decision*), causative uses (e.g., make something possible), earn money (e.g., *make a living*), link verb uses (e.g., *she will make a good teacher*), make it (idiomatic) (e.g., *if we run, you should make it*), phrasal/prepositional uses (e.g., *make out*), other conventional uses (e.g., *make one's way*). (p. 177).

The researchers categorized every use of the verb *make* in the three corpora based on the categories of use listed above. The results of this categorization indicated that the three corpora are similar in terms of the ranking order of the uses. The causative uses are the most common in all the three corpora, followed by the delexical uses while the other uses are apparently less common.

Regarding the frequencies of the different uses, the results indicated that the delexical uses are significantly less frequent in FR and SW than in LOCNESS. Altenberg and Granger (2001) claimed that, "Learners may at the same time overuse a high-frequency verb and underuse its delexical structures" (p. 178). While a third of the uses of the verb *make* in LOCNESS used with nouns expressing speech actions (e.g., *argument, claim*, *point*), only 9 to 13% of the learners' uses do. Altenberg and Granger (2001) pointed out that EFL learners do not only underuse delexical structures, but also misuse them.

Furthermore, regarding the frequency of the causative structures, the results indicated that causative uses are more frequent in SW than in FR and LOCNESS. In a detailed analysis, Altenberg and Granger (2001) described the causative *make* constructions as follows:

- Adjective structure: *make* + object + adjective (e.g., *make something possible*).

- Verb structure: *make* + object + verb (e.g., *make somebody realize something*).

- Noun structure: *make* + object + noun (e.g., *make somebody a star*).

The researchers' goal was to investigate the distribution of these constructions in the three corpora. The results revealed that Swedish EFL learners overuse adjective and verb structure compared to French EFL learners and native speakers. On the other hand, French-speaking learners underuse causative *make* structures, adjective and noun structures in particular. Altenberg and Granger (2001) attributed the striking differences between French- and Swedish-speaking learners in using the high-frequency verb *make* to interlingual, intralingual, and inadequate teaching forces.

From the results of their study, Altenberg and Granger (2001) concluded that although some verbs (such as *make*) are high frequently used in English, they can be problematic for some EFL learners. These verbs tend to be neglected after being taught although they are very complex ones. For example, the delexical uses of the verb *make* seem to be particularly error-prone. However, even the causative uses that seem extremely safe are found problematic, too in this study.

<u>Yoo (2009)</u>

This study sought to compare what ESL/EFL grammar books say about the definite article *the* with what a corpus-based analysis indicates about *the* usage. Yoo (2009) used six popular grammar series with separate volumes and for different levels (a total of 21 books). For the corpus findings, Yoo (2009) used a corpus-based analysis of data taken from the *Longman Grammar of Spoken and Written English.* This corpus included language used in four major registers: conversation, fiction, newspaper, and academic prose and in two other additional registers that are non-conversational speech and general prose.

Yoo (2009) presented a brief overview of the literature on the definite article *the* before starting to compare its usage/coverage in ESL/EFL grammar books and in the corpus. He categorized *the* uses into three main groups as following:

1. Referential uses:
   a. Anaphoric use.
   b. Associative use.
   c. Situational use.
   d. Cataphoric use.
   e. Unique reference.
   f. Sporadic reference.
2. Generic uses.
3. Non-referential uses.

In addition, the researcher differentiated between the definite article *the* and the "most definite" null article, stating that two forms of null (zero) article exist: the zero article $\varnothing_1$ precedes non-count nouns and plural count nouns (e.g., Athena loves $\varnothing_1$ milk and $\varnothing_1$ cookies) while the null article $\varnothing_2$ precedes singular proper nouns and some singular count nouns (e.g., $\varnothing_2$ Matthew was $\varnothing_2$ best man at my wedding). (Yoo, 2009, p.269).

The results of Yoo's (2009) comparison indicated that ESL/EFL grammars cover the three most important referential uses of the definite article *the*: second mention (anaphoric use), shared knowledge (under which associative use subsume), and situational uses. Most ESL/EFL grammars are extremely good in explaining second mention and giving examples of it. However, none of these grammar series explain and give examples of shared knowledge and situational uses separately. They do not provide separate explanation for shared knowledge and situational uses, but rather they describe them in a way that requires the learners themselves to discover when a noun is definite. The results also indicated that ESL/EFL grammars intentionally avoid discussing the generic uses of *the*, especially for beginning students. Likewise, the null article ($\varnothing_2$) is not given attention in ESL/EFL grammar series.

The corpus findings, on the other hand, indicated that the situational use of *the* is the most common in conversations. However, ESL/EFL material designers believe that anaphoric use of *the* is the most common, so they give it more attention in their books. In addition, ESL/EFL material writers pay no attention to cataphoric uses of the definite article *the* although the corpus findings indicated that cataphoric uses should be discussed in detail in ESL/EFL materials (especially for advanced students) since they are frequently found in newspaper language and academic prose.

<u>Narita, Sato, and Sugiura (2004)</u>

In this study, Narita, Sato, and Sugiura (2004) selected 25 logical connectors to compare their use in essays written by advanced university Japanese students and university NSs. Also, Narita et al. (2004) investigated how Japanese students differ from other NNSs (advanced French, Swedish, and Chinese learners) in their use of these connectors. The 25 connectors examined in this study were categorized into four semantic categories: enumeration/addition (e.g., *first, next, in addition, also*), apposition (e.g., *for example, for instance*), result/inference (e.g., *therefore, thus, as a result*), and contrast/concession (e.g., *on the other hand, in contrast, however*).

Narita et al. (2004) used two sub-corpora of the International Corpus of Learner English (ICLE): the Japanese component of the ICLE corpus (referred to as JPICLE) and the Louvain Corpus of Native English Essays (referred to as LOCNESS-US). The data was analysed not only to investigate the frequency rates of the 25 connectors, but also to examine their occurrence positions in every instance. The researchers extracted every instance of the logical connectors with its adjacent context. Then, they annotated every connector's occurrence position as I (Sentence-Initial), M (Middle), and F (Sentence-Final). In the next step, Narita et al. (2004) used computations to count the frequency rate of each connector per its occurrence position.

The results of this study indicated that, compared to NSs' of the 25 logical connectors, Japanese students tend to overuse enumerative/additive, appositive, and resultative connectors. However, they tend to underuse inferential and contrastive connectors. In fact, Japanese students overused five connectors: *for example, of course, first, moreover,* and *in addition* while they underused *then, yet,* and *instead*. Furthermore, the study shows that Japanese students and NSs

are different in their positional tendencies. While NSs prefer to use connectors in the sentence-initial position and in the middle position as well, Japanese students seem to have a strong preference in using connectors positioning mostly in the sentence-initial. Narita et al. (2004) attributed NNSs' tendency to use logical connectors in the sentence-initial position to their attempt to show cohesion between sentences.

In addition, the research findings indicated that there exist some similarities and differences between the four groups of NNSs in using logical connectors. For example, all NNSs tend to overuse some logical connectors as *moreover, for example,* and *for instance*. Also, NNSs, except Chinese learners, over use the resultative connector *of course*. Likewise, NNSs share some similarities in underusing some logical connectors including *yet* and *instead*. Narita et al. (2004) attributed NNSs' underuse of such connectors to their low familiarity with the use of these connectors.

Hinkel (2001)

This study examined the way explicit cohesive devices were used in academic essays written by students speaking English, Japanese, Korean, Indonesian, and Arabic as their mother tongue. Hinkel (2001), in her study, focused on the median frequency rates of uses of phrasal-level coordinators (e.g., *also, both, either…or, neither…nor*), sentence transitions (e.g., *first(-ly), second(-ly), in addition, therefore, hence, besides*), logical-semantic conjunctions (e.g., *as well, because of*), demonstrative pronouns (e.g., *this, that*), and enumerative (e.g., *advantage, aspect*) and resultative nouns (e.g., *finish, effect*).

Hinkel (2001) used a corpus of 897 academic written texts (a total of 265,812 words) produced by 1,101 participants (895 NNSs and 206 NSs). The essays were written in a 50-minutes period in response to one of the following prompts:

1. Some people believe that when parents make their children's lives too easy, they can actually harm their children instead. Explain your views on this issue. Use detailed reasons and examples.

2. Many people believe that grades do not encourage learning. Do you agree or disagree with this opinion? Be sure to explain your answer using specific reasons and examples.

3. Some people learn best when a classroom lesson is presented in a serious, formal manner. Others prefer a lesson that is enjoyable and entertaining. Explain your views on this issue. Use detailed reasons and examples.

4. Many educators believe that parents should help to form children's opinions. Others feel that children should be allowed to develop their own opinions. Explain your views on this issue. Use detailed reasons and examples.

5. Some people choose their major field of study based on their personal interests and are less concerned about future employment possibilities. Other choose majors in fields with a large number of jobs and options for employment. What position do you support? Use detailed reasons and examples. (Hinkel, 2001, p. 118-119).

To analyze the data, Hinkel (2001) counted the number of words in each essay and the number of occurrences of each of the overt cohesive devices by hand. Then, she performed computations to identify percentage rates of the cohesive devices.

The results of the study indicated that even advanced NNSs rely on restricted features in constructing cohesive texts. The NSs' and NNSs' use of explicit cohesive devices is as follows:

- Phrasal-level coordinators: compared to the median frequency rates of using coordinating conjunctions in NSs' texts, NNSs speaking Japanese and Korean utilize phrasal-level coordinators in rates relative to NSs'. However, Indonesian students employ fewer coordinators while Arabic students employ them even more frequently than NSs.

- Sentence transitions: all groups of NNSs use sentence transitions significantly more frequent than NSs. NNSs tend to overuse sentence transitions as their most predominant explicit means of joining their ideas together, even when their ideas are seemingly irrelevant.

- Logical-semantic conjunctions: NSs and NNSs are similar in their use of logical-semantic conjunctions. This kind of cohesive devices is not very common in their texts.

- Demonstrative pronouns: The use of demonstrative pronouns in the texts of all groups NNSs, except Indonesians, exceeds NSs' use.

- Enumerative and resultative nouns: Korean students use enumerative nouns more frequently than NSs, but Indonesian and Arabic students use resultative nouns more frequently than NSs. In general, the use of enumerative and resultative nouns is rear in the texts of all participants.

Lu (2011)

In this study, Lu (2011) evaluated 14 syntactic complexity measures to determine ESL writers' written language development. The researcher described syntactic complexity as

36

"evident in second language variation and sophistication, or, more specifically, the range of syntactic structures that are produced and the degree of sophistication of such structures" (Lu, 2011, p.36). Lu (2011) classified the 14 syntactic complexity measures evaluated in this study into five types as following:

Type 1: Length of production

1. Mean length of clause (MLC).

2. Mean length of sentence (MLS).

3. Mean length of T-unit (MLT).

Type 2: sentence complexity

4. Clause per sentence (C/S).

Type 3: subordination

5. Clause per T-unit (C/T).

6. Complex T-units per T-unit (CT/T).

7. Dependent clauses per clause (DC/C).

8. Dependent clauses per T-unit (DC/T).

Type 4: coordination

9. Coordinate phrases per clause (CP/C).

10. Coordinate phrases per T-unit (CP/T).

11. T-units per sentence (T/S).

Type 5: particular structures

12. Complex nominals per clause (CN/C).

13. Complex nominals per T-units (CN/T).

14. Verb phrases per T-unit (VP/T).

Lu (2011) used these syntactic complexity measures to analyze large-scale data collected from the Written English Corpus of Chinese Learners (WECCL), a corpus that consisted of 3,678 essays written by ESL learners from nine Chinese colleges. The corpus was analyzed using a computational system, designed to automate the measurement of syntactic complexity of the writing samples produced by ESL college-level writes. This analysis enabled Lu (2011) to answer four major research questions:

1. What is the impact of sampling condition, including institution, genre, and timing condition, on the mean values of any given syntactic complexity measure?

2. Which measures show significant between-proficiency differences? What is the magnitude at which between-proficiency differences in each measure reach statistical differences?

3. What are the patterns of development for the measures that show significant between-proficiency differences?

4. What is the strength of the relationship between different pairs of syntactic complexity measures? (Lu, 2011, p. 46).

In this study, the participants were from nine different institutions, and they wrote in different genres (i.e., argumentative and narrative) in different timing conditions (i.e., timed and untimed). Lu (2011) found that institution, genre and timing condition significantly impact the

relationship between syntactic complexity and proficiency. Argumentative and untimed essays showed higher syntactic complexity than narrative and timed essays.

The second research question sought the syntactic complexity measures that show significant between developmental levels. From all the 14 measures, Lu (2011) found only 10 measures discriminate between proficiency levels, including MLC, MLS, MLT, CS, DC/C, DC/T, CP/C, CP/T, CN/C, and CN/T. The best candidates that indicate development between levels are categorized into three groups: measure that discriminate two or more adjacent levels and increase linearly across all four levels (CN/C and MLC), measures that discriminate two adjacent levels and progress in a way relevant to school level (CN/T, MLS, and MLT), and measures that discriminate nonadjacent levels and increase significantly from lower to higher levels (CP/C and CP/T). (Lu, 2011, p. 53).

The results, in addition indicated several patterns of development for the 10 measures mentioned above. Lu (2011) grouped these 10 measures, first, according to whether the changes are positive or negative and, second, according to whether or not the pattern of development is linear across the levels. The researcher found that seven measures out of ten show positive observed changes form lower to higher levels and increase linearly across the four levels, including MLC, MLS, MLT, CP/C, CP/T, CN/C, and CN/T. On the other hand, three measures show negative changes from lower to higher levels and progress nonlinearly, including C/S, DC/C, and DC/T.

The results of this study also indicated the correlation between the 14 syntactic complexity measures. It was found that measures correlate strongly with other measures of the

same type or that of the same structure (Lu, 2011, p. 53). For example, high correlations were found between MLS and MLT, C/T and DC/T, CN/C and CN/T, and between CP/C and CP/T.

<u>Hunston and Gill (1998)</u>

This study described the verbs defined in the Collins COBUILD English Dictionary (CCED), investigating the complementation patterns of each verb. The corpus used in this study was the 250-million-word Bank of English at Collins and Birmingham University International Language Database (COBUILD).

The researchers used the concordance lines for each verb, arranged based on the type of group, phrase, clause, that follows the verb. Hunston and Gill's (1998) major goal was to provide symbols that can represent the patterns of groups of similar concordance lines. For example, for the verb *consider*, Hunston and Gill (1998) found different patterns: 'V n to-inf' (i.e., the verb is followed by a noun group and a to-infinitive clause), '*be* V-ed to-inf', (i.e., the passive form of the verb followed by a to-infinitive clause), 'V n n' (i.e., the verb is followed by two noun groups), '*be* V-ed n' (i.e., the passive form of the verb followed by a noun group), 'V n adj' (i.e., the verb is followed by a noun group and an adjective group), '*be* V adj' (i.e., the passive form of the verb is followed by an adjective group), etc. Through the concordance lines, there were found three main senses of the verb *consider* that were already distinguished by CCED: to have an opinion about, to think about, and possibly intend to do something. (Hunston & Gill, 1998, p. 48).

The results of this study also indicated that verbs that share a pattern share also meaning. For example, the researchers provided a list of 48 verbs that share the pattern 'V *as* n' (i.e., the

verb, or verb and particle, is followed by a prepositional phrase beginning with *as*). Most verbs are found close to synonyms, so it is possible to divide these verbs into meaning groups. For instance, verbs concerned with having a role or function (*act, double, double up, figure, function, operate, serve*), verbs concerned with beginning and ending: beginning (*begin, originate, start, start off, start out*) continuing (*continue, remain*) and ending (*end, end up, finish, finish up*), and verbs concerned with being perceived in a particular way (*come across, come over, count, emerge, go down, pass, qualify, rank, rate, shape up*). Likewise, other verbs that share a pattern can have a similar division of meaning groups. This study pointed out that syntax and lexis are interdependent areas of language study, and that ELLs can learn which verbs have which patterns by learning different meaning groups that consist of verbs of a shared pattern. (Hunston & Gill, 1998, p. 62).

Alzuhairy (2016)

The aim of this thesis was to investigate the most frequent verb tenses used by NSs in their writings. The corpus used in this thesis was a collection of 40 research papers (consisted of over 100,000 words) written by undergraduate NS students. The papers were analyzed using a code coloring method to facilitate eliciting the various targets (twelve verb tenses [simple present, simple past, present perfect, present progressive, past progressive, simple future, future progressive, past perfect, future perfect, present perfect progressive, past perfect progressive, future perfect progressive], modals, perfect modals, and imperatives). Then, Excel document files were used to calculate the median frequency rates for each target in each paper.

The results revealed that the simple present tense was the most predominant verb tense, representing the highest percentage (49.99%). Simple past came next with a percentage of 28.50% and present perfect was the third most frequently used tense with a percentage of only 4.65%.

Comparing the verb tenses based on aspect, there existed significant differences between the tenses within the simple aspect (i.e., simple present 49.99%, simple past 28.05%, simple future 2.32). Tenses within the progressive aspect were not very common on NSs' writings. The most frequently used progressive tense was the present progressive with a percentage of 2.31%. however, the least common progressive tense was the future progressive (0.02%). In addition, tenses within the perfect aspect showed striking differences. Although the present perfect was ranked as the third most common verb tense used by NS writers, the future perfect constituted only 0.01%. Finally, among the three tenses within the perfect progressive aspect, the present perfect progressive represented the highest percentage (0.219%) while the other two tenses (i.e., the past perfect progressive and the future perfect progressive) were the least frequently used verb tenses, representing only 0.009%. However, all three were extremely infrequent.

The findings also revealed the NSs' use of the three non-tense categories, namely modals, perfect modals, and imperatives. Interestingly, it appeared that NSs prefer modals (9.9%) over perfect modals (0.3%) and imperatives (0.2%).

Researching Another University Corpus

Alzuhairy (2016) examined the median frequency rates of the twelve verb tenses and three other categories as used by NS students in their academic written composition. The simple

present tense dominated in his corpus of 103,181 words, constituting half of the entire corpus with a percentage of 49.99%. The corpus Alzuhairy (2016) used was a collection of assignments written by students who were taking general composition courses in the Department of Writing and Rhetoric, University of Central Florida. In addition, the papers contained controversial topics that Alzuhairy (2016) believed to have a variety of tense usages.

While Alzuhairy's (2016) results are interesting, we still know relatively little about writing done for other General Educational Program (GEP) courses. It has come to the researcher's attention that widening research on verb tenses usage by NSs in their writings would help to draw general conclusions about the frequency of verb tenses and help to enrich and foster TESOL educators' knowledge about this particular grammar aspect. In fact, history is a substantial subject that almost all university students need to take. However, it is a subject that no study has ever examined verb tenses in its composition. In this study, the aim is to investigate the frequency of verb tenses in composition written by NSs for a history class. Therefore, the current study seeks to gain data on the usage of the same 15 categories within a different academic subject.

CHAPTER 3: METHODOLOGY

Introduction

As verb tenses usage is considered a substantial grammar issue for ESL/EFL educators and leaners, an investigation of the actual, normal usage of the twelve verb tenses by NSs is crucial. The fact that TESOL educators as well as ESL/EFL learners need to know is that some verb tenses are much more prevalent than the others. Therefore, this research study has been conducted for the sake of answering the following research question:

- What are the most frequently used verb tenses that native writers use in assignments in university history courses?

Design of the Study

This is a corpus-based study. The study was based on a corpus that consisted of more than 100,000 words, gathered from NSs' written assignments for a history class. The collected assignments were a total of 130 papers written by 65 undergraduate NS students. Each paper was analyzed to identify uses of the twelve verb tenses and the three other verb forms (modals, perfect modals, and imperatives) as research targets. To calculate the frequency rates, Microsoft Excel was used, giving the total numbers.

The Corpus of the Study

The whole corpus was gathered from the Department of History in the University of Central Florida (UCF). The corpus, in its essence, is a collection of assignments that were written for the history course: Western Civilization, a course that almost all undergraduate students are

required to take. The corpus consisted of 101,713 words, collected from 130 assignments written by 65 students (i.e., two assignments per student).

Procedure

A professor from the Department of History provided a massive number of assignments that were written for the Western Civilization class by undergraduate NS students. The researcher randomly chose 130 papers, all written by native writers, whose word count is large enough to create an authentic corpus of more than 100,000 words (i.e., 101,713 words). All the papers were coded with numbers from 001 to 130. Furthermore, some parts such as bibliographies and footnotes were removed from the students' papers since the researcher is merely interested in the main content of the papers. For privacy protection, personal information, including names and dates, were removed, too. After removing all the unrelated parts from the students' papers, the researcher is left with the main content (the corpus) that she aims to use for the purpose of this study. The papers are 2-4 pages in length and range between 407 and 1263 words in terms of word count. This indicates, the researcher assumed, that the students differ in their writing abilities to convey their thoughts and ideas.

Analysis

The procedure of analyzing the data was straightforward. After removing all the unrelated parts from the students' papers, the researcher was ready to analyze the data. She used a code-coloring method to label the twelve verb tenses and the three other verb forms (modals, perfect modals, and imperatives) as shown in the following table (Table 3). It is worth mentioning that although the researcher thinks that the code-coloring system she used was

extremely useful for the purposes of this study and for readers who are viewing this research

report on a screen with color capability, people without access to color capability may not be

able to understand the code-coloring system employed in this study.

Table 3: Code-Coloring System Used to Highlight the Targets

Target	Code-Coloring System
Simple present	yellow
Present progressive	turquoise
Present perfect	green
Present perfect progressive	pink
Simple past	blue
Past progressive	violet
Past perfect	red
Past perfect progressive	Dark green
Simple future	Dark yellow
Future progressive	Light gray
Future perfect	teal
Future perfect progressive	Dark blue
Modals	gray
Perfect modals	Dark red
Imperatives	**Bold**

To verify the researchers' accuracy of eliciting and labeling the targets, 10% of the

corpus (i.e., 13 papers including 9,998 words) was selected randomly and examined by an

independent reviewer. The 13 randomly selected data-papers are: 003, 018, 021, 026, 038, 044,

055, 062, 079, 082, 091, 119, and 126. A comparison of the researcher's labeling and the

independent reviewer labelling showed the findings were identical. They both found that these

13 papers contain 985 verb tenses (*simple present= 398, present progressive= 20, present*

perfect= 19, present perfect progressive= 0, simple past= 485, past progressive= 27, past

perfect= 33, past perfect progressive= 0, simple future= 3, future progressive= 0, future

perfect= 0, future perfect progressive= 0), 84 modals, 10 perfect modals, and 2 imperatives. This

confirmed the reliability of the researcher's findings and verified the accuracy of her process of analyzing the data.

While analyzing the data, the researcher, as well as the independent reviewer, who completed only 10% of the data, encountered different issues about some contexts that worth mentioning.

- Sentences written in a passive voice, questions, and negative forms of sentences were all treated the same way as affirmative sentences. The only thing mattered was the main targets (the twelve verb tenses, modals, perfect modals, and imperatives).

- Imperatives frankly have no tense, so they were put in a different category. They were not marked with a color, but put in bold instead.

- When some students committed spelling mistakes, the correct word the student intended to use was considered. For example, in paper 044, the student used 'were' instead of 'where', so the mistakenly used verb (i.e., 'were') here was not marked. However, in other papers, it was found that some students wrote 'would of been' instead of 'would have been'. In this case, a use of a perfect model was considered and it was marked with a dark red color.

- When it was encountered that some students mistakenly used two verbs in one sentence, only one was considered and highlighted.

- 'Will' was always considered a simple future marker unless it indicated another usage (e.g., asking for permission), then it was marked as a modal/ perfect modal.

Once the entire data was analyzed, all the occurrences of verb tenses, modals, perfect modals, and imperatives were counted and the numbers were inserted in Microsoft Excel, which then automatically calculated the numbers and gave the totals. Each target was placed in a row followed by 130 columns each represented a research paper. The researcher had the number of all the occurrences of the targets in the corpus, and with the use of Microsoft Excel, she had the numbers of each target added together, showing the frequency of the targets.

CHAPTER 4: FINDINGS

Introduction

In the teaching of English to non-native speakers, the use of verb tenses has been considered a crucial area of grammar. As a result, authors of grammar books always include multiple chapters on verb tenses, in addition to explanations of their forms and different uses, followed by numerous activities, exercises, and practices that would reinforce ESL/EFL learners' comprehension of English verb tenses. In these books, each verb tense has its own section, and sometimes chapter, in the book. Each verb tense tends to receive the same amount of instruction, but is equal treatment appropriate?

Allocating equivalent amounts of pages for each verb tense assumes that each is needed in equivalent proportions by ESL/EFL learners. However, our intuition tells us that some tenses such as simple past tense (e.g., she went) are used more frequently than some others such as future perfect tense (e.g., she will have gone). The purpose of the present study is to measure the relative frequency of the twelve verb tenses and three other forms, namely, modals, perfect modals, and imperatives.

In investigating the frequency of the twelve verb tenses, the current study answered the following question:

- What are the most frequently used verb tenses that native writers use in assignments in university history courses?

Results

To answer this research question, the researcher created a corpus of 101,713 words

gathered from 130 assignments written by undergraduate NS students for their history course:

Western Civilization. This corpus was analyzed to find the relative frequency of the twelve verb

tenses, modals, perfect modals, and imperatives.

The following table (Table 4) provides the number of occurrences and the percentages of

all verb tenses used by NS undergraduate students in their academic history composition. The

table organizes the verb tenses from the most frequent verb tense to the least.

Table 4: The Number of Occurrences and Percentages of the Twelve Verb Tenses Elicited from a 101,713-Word Corpus

VERB TENSE	NUMBER OF OCCURRENCES	PERCENTAGE
simple past	5002	44.34%
Simple present	4440	39.35%
Past perfect	183	1.62%
Past progressive	150	1.33%
Present perfect	138	1.22%
Present progressive	120	1.06%
Simple future	93	0.82%
Past perfect progressive	5	0.04%
Present perfect progressive	3	0.03%
Future progressive	1	0.008%
Future perfect	1	0.008%
Future perfect progressive	0	0%

To support the corpus study, other linguistic targets besides the twelve verb tenses were

included: modals, perfect modals, and imperatives. Table 5 shows the number of occurrences and

percentages of these three verb groups.

Table 5: The Number of Occurrences and Percentages of Modals, Perfect Modals, and Imperatives

	NUMBER OF OCCURRENCES	PERCENTAGE
Modals	970	8.6%
Perfect modals	155	1.37%
Imperatives	21	0.19%

In this corpus-based study, the simple past tense was dominant, representing the highest percentage (44.34%). The simple past tense occurred 5,002 times and was used extensively to recount past events as the papers were discussing stories and issues that happened in the past.

The following is an example:

Excerpt A:

"The two rulers Constantine and Clovis both *came* to Christianity by being in the same predicament. Both the rulers *were* in a war with another country". (Excerpted from data-paper 045).

The verb tense that had the second highest percentage was the simple present tense (39.35%). It was mostly used wherever a student wants to state a fact. For example,

Excerpt B:

"The excerpt from the Arabian Chronicler *doesn't give* any details about the actual battle". (Excerpted from data-paper 021).

The simple present tense is also used in the opening paragraph(s) and sometimes in the last paragraph(s) as well, especially when referring to famous people and books. For example,

Excerpt C:

"Also, the St. Denis Chronicle *gives* us the reasoning behind why the Muslims are invading". (Excerpted from data-paper 021).

Excerpt D:

"Montesquieu's document *stresses* the importance of a balanced government with branches that *are* separate in order to prevent tyranny and corruption. Montesquieu *states*, "There would be an end…" (Excerpted from data-paper 089).

As shown above, simple present and simple past were the highest frequently used tenses in the corpus, with seemingly more occurrences of the simple past tense over the simple present. What comes a third in order is the past perfect tense with a percentage of only 1.62%. That is, 183 occurrences. The third most common verb tense in this corpus was found to occur much less frequently than modals, which represented a percentage of 8.6% (970 occurrences).

When comparing verb tenses based on tense (i.e., present, past, future), past verb tenses (simple, progressive, perfect, and perfect progressive) occurred more frequently than present and future verb tenses, as shown in the following chart (Figure 1).
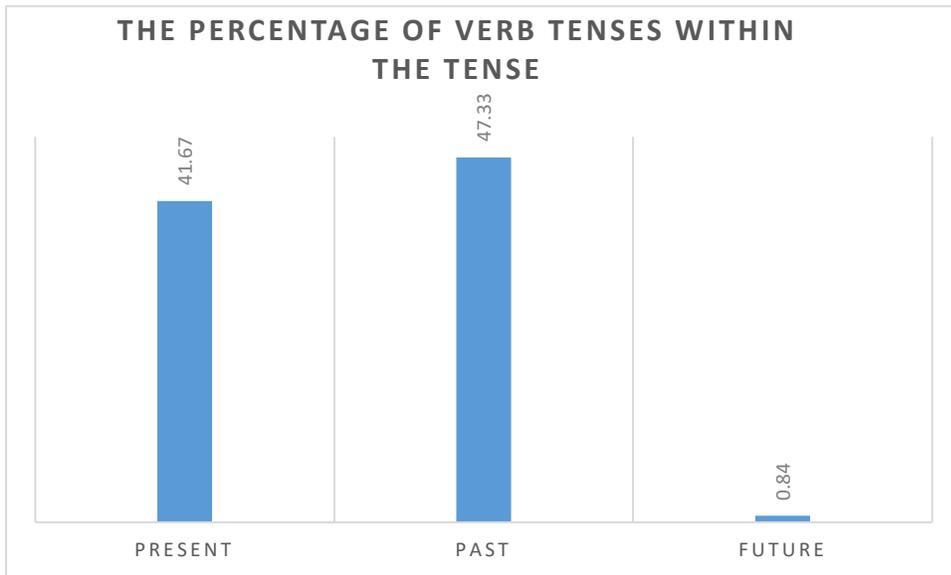
Figure 1: The Percentage of Verb Tenses Within the Tense

The results also showed a comparison between the frequency of verb tenses within the same aspect (simple, progressive, perfect, and perfect progressive). The following chart (Figure 2) depicts the occurrences of verb tenses within the simple aspect. The chart shows the drastic low use of the simple future compared to the simple present and past.

THE PERCENTAGE OF VERB TENSES WITHIN THE SIMPLE ASPECT

PRESENT	PAST	FUTURE
39.35	44.34	0.82
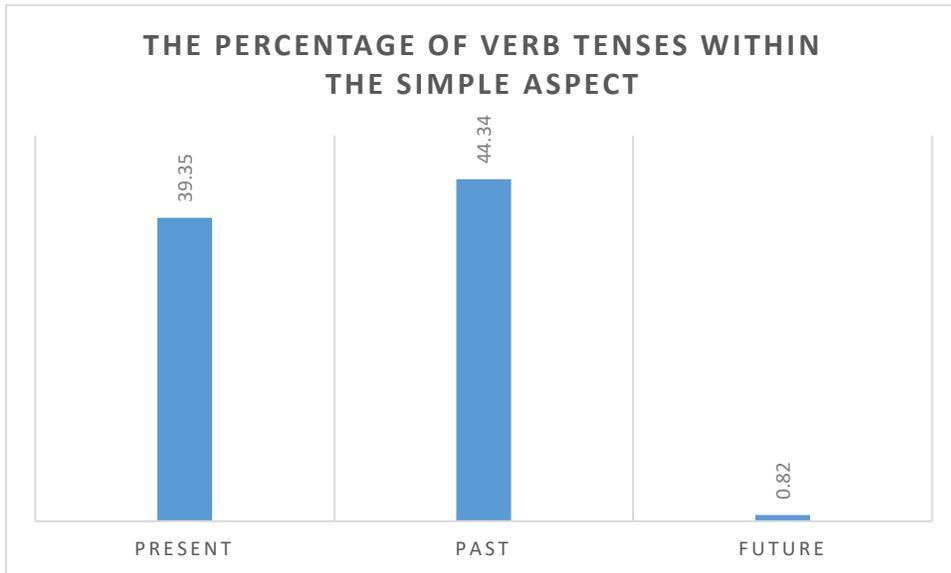
Figure 2: The Percentages of Simple Verb Tenses

The following chart (Figure 3) shows the percentage of verb tenses within the progressive aspect. The past progressive tense has the highest percentage. However, the difference between the percentage of the present progressive (120 occurrences) and the past progressive (150 occurrences) is not striking. The future progressive tense was used only once.
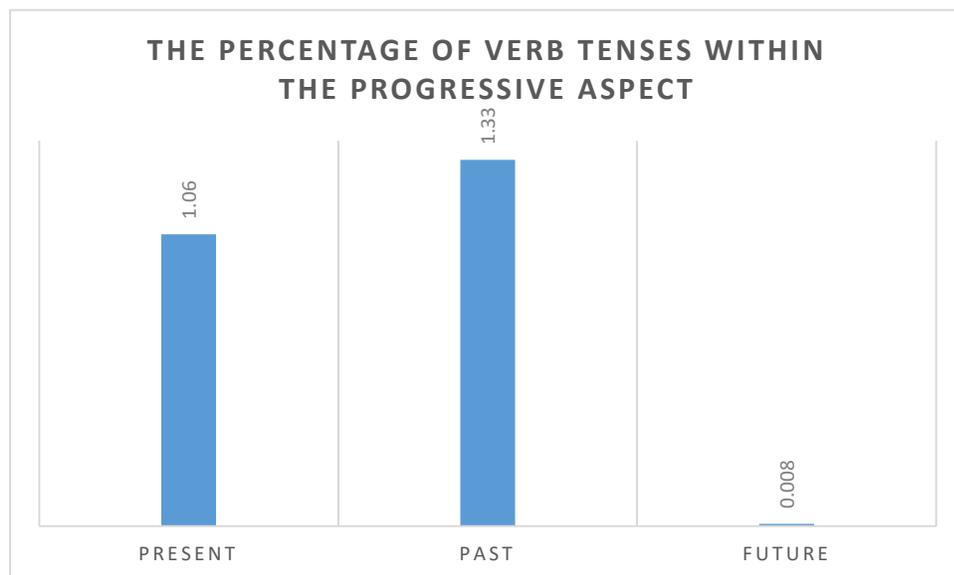
Figure *3*: The Percentages of Progressive Verb Tenses

It was noticed that out of the 150 occurrences of the past progressive tense, there were six occurrences in the form of *was/were + going + to* as in the following examples:

Excerpt E:

"This led him to believe that there was a God and that they *were going to* win the battle." (Excerpted from data-paper 045).

Except F:

"Martin *was going* to leave his wife, kids, and family forever but when he heard of Arnaud taking his place he came back". (Excerpted from data-paper 011).

The results found 138 uses of the present perfect tense, 183 uses of the past perfect tense, and only one use of the future perfect tense. The percentage of the tenses within the perfect tense are shown in Figure 4.

THE PERCENTAGE OF VERB TENSES WITHIN THE PERFECT ASPECT

1.22 PRESENT

1.62 PAST

0.008 FUTURE
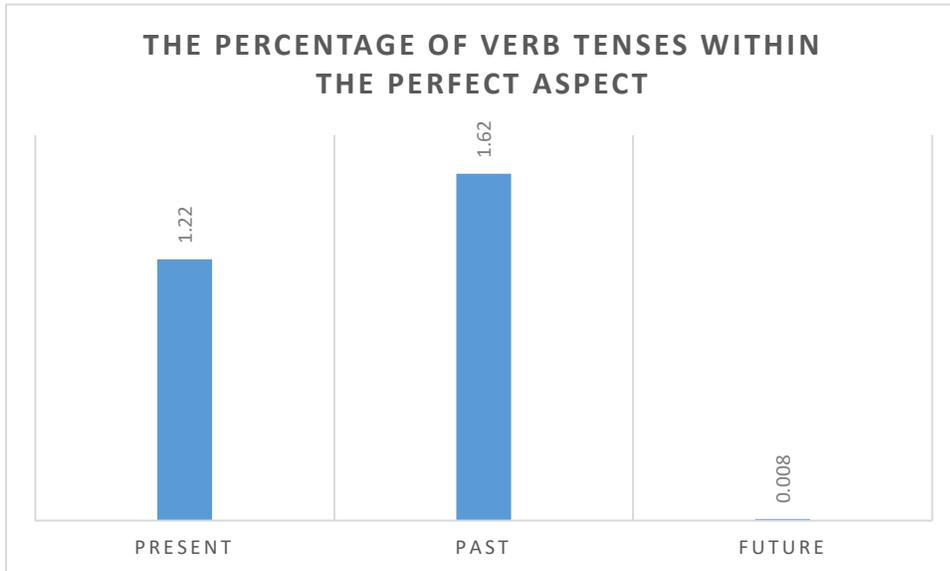
Figure 4: The Percentages of Perfect Verb Tenses

The results found extremely few occurrences of the verb tenses within the perfect progressive tense: 3 uses of the present perfect progressive, 5 uses of the past progressive tense, and the future perfect progressive was never used in this corpus. Figure 5 shows the percentages.

THE PERCENTAGE OF VERB TENSES WITHIN
THE PERFECT PROGRESSIVE TENSE

0.04

0.03

0

PRESENT   PAST   FUTURE
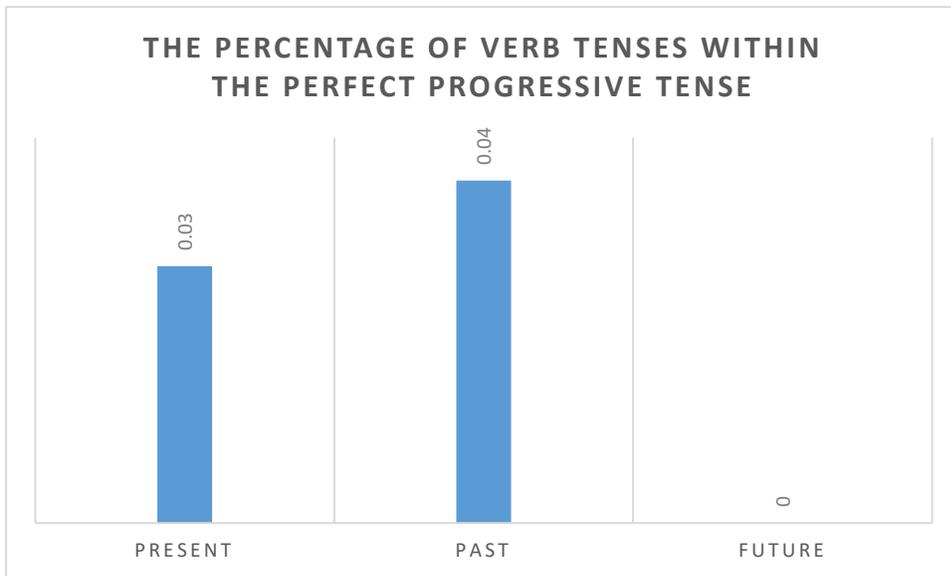
Figure 5: The Percentages of Perfect Progressive Tenses

As the results revealed, modals were used as frequently as 8.6% in the corpus. Paying a close attention to the native writer's use of modals in their writing, the researcher found that the modal *would* was the most predominant, followed by the modal *can*. Table 6 shows the number of occurrences of each modal used in this corpus and orders them from the most frequently used modal to the least.

Table 6: The Occurrences of Modal Verbs in the Corpus

MODAL	NUMBER OF OCCURRENCES OUT OF 970	PERCENTAGE
Would	295	30.41%
Can	223	22.99%
Could	175	18.04%
Should	103	10.61%
May	45	4.03%
Must	38	3.92%
Be able to	37	3.81%
Might	29	2.99%
Have to	21	2.16%
Ought to	2	0.20%
Shall	1	0.10%
Be supposed to	1	0.10%

CHAPTER 5: CONCLUSIONS

Introduction

The current study was conducted to determine the relative frequency of the twelve verb tenses and the three other verb forms (modals, perfect modals, and imperatives) in a corpus of 101,713 words written by undergraduate NS students for their history class. The study was implemented to answer the following question:

- What are the most frequently used verb tenses that native writers use in assignments in university history courses?

Conclusions

The results of this corpus-based study revealed that the simple past tense was the most frequently used verb tense among all verb tenses, with 5,002 occurrences (44.34%). The simple present tenses came next with a frequency of (39.35%). Modals were with third position, occurring more frequently than all the other linguistic items targeted in this study together, with a percentage of (8.6%).

Each of the remaining verb items occurred very rarely. The order of the remaining language targets was as follows: past perfect (1.62%), perfect modals (1.37%), past progressive (1.33%), present perfect (1.22%), present progressive (1.06%), and the rest of the verb tenses and imperatives were lower than 1%.

While the current study focused only on history papers, previous research looked at student writing for university English courses. Alzuhairy (2016) assembled a corpus of 103,181

59

words from papers written by undergraduate NS students for general composition courses. The following table (Table 7) compared the frequency of verb tenses in these two different writing corpora: general composition and history.

Table 7: The Percentages of Verb Tenses in Two Corpora

VERB TENSE	COMPOSITION	HISTORY
Simple present	49.99%	39.35%
Present progressive	2.31%	1.06%
Present perfect	4.65%	1.22%
Present perfect progressive	0.21%	0.03%
Simple past	28.50%	44.34%
Past progressive	0.71%	1.33%
Past perfect	0.73%	1.62%
Past perfect progressive	0.009%	0.04%
Simple future	2.32%	0.82%
Future progressive	0.02%	0.008%
Future perfect	0.01%	0.008%
Future perfect progressive	0.009%	0%
Modals	9.9%	8.6%
Perfect modals	0.3%	1.37%
Imperatives	0.2%	0.19%

Each of the two studies used a corpus from a different field, but they mostly found similar information. They revealed a prevalence of the simple present and simple past tense, with modals in third position. In the current study, simple past was first and simple present second, but the reverse was found in Alzuhairy's (2016) study. Modals were used at similar rates, with 9.9% for the English papers and 8.6% for the history papers.

The frequency of present tense, past tense, and modals was 88.39% in the English papers and 92.90% in the history papers. Therefore, these three categories account for approximately 90% of each subject area.

What is interesting from these findings is that the verb tenses observed in this study do not occur with equivalent frequency. On the contrary, three categories dominate with a combined total of approximately 90%, while the other twelve categories share approximately 10% frequency.

The frequency of a verb tense or any other aspect of a language, e.g., the direct article *the* or a gerund, can surely inform language curriculum, but a lower frequency for a verb tense does not necessarily mean it should not be taught in the curriculum. For instance, the present perfect tense was not found very frequent in the current study, occurring in only 1.22% of main verbs. While the most common meaning of present perfect may be the past action that is still true now (e.g., The U.S. has been an independent country for more than two hundred years), this verb tense is an important verb tense in academic writing because it serves to connect a past action that is relevant to the present (Folse, 2016), as seen in the following four examples from student papers in this study:

- To emphasize an event whose consequences are still contentious or important to the current time.

Excerpt G:

"Lifestyle and society *has changed* tremendously over the past 500 years with some being for the better and others for worse." (Excerpted from data-paper 044).

- To indicate what has been previously found/discussed.

Excerpt H:

"History *has already showed* us that these were tough times." (Excerpted from data-paper 037).

- To allow the writer to transition from one topic to another or from one facet of a topic to another facet of that same topic

Excerpt I:

Government is essential to ensuring the natural rights of humans, however in the past there *have been* different theories about how a government should be run, what type it should be, where its authority comes from, and basically what constructs it. (Excerpted from data-paper 050).

- To talk about past experiences without mentioning a specific time such as *yesterday* or *2015*

Excerpt J:

"Musa, the leader of the Arab forces, declares that his army *has never been beaten*." (Excerpted from data-paper 043).

Implications

For Teachers

The findings of this study can improve the way ESL/EFL teachers teach verb tenses to their learners. The results indicated the most frequently used verb tenses, so TESOL educators can now be informed about which tenses should be given more time and effort. By a large gap,

simple past and simple present were found the most frequent verb tenses while future

progressive, future perfect, and future perfect progressive were found to be the least frequently

used verb tenses. This does not mean that these verb tenses should not be taught in ESL/EFL

classrooms, but it might imply that they should not be given as much attention as they often are.

For Material Designers

Materials designers, especially those of grammar books, should inform students of the

strong prevalence of two verb tenses over all the others, especially in academic writing. This

could help the students to focus, first, on what they really need to learn, and not get confused or

distracted by the various verb tenses of English. Furthermore, the results can help material

designers to design their grammar books in a way that gives more attention to the most frequent

verb tenses through explanations, practices, and exercises. Material writers can depend on the

results of this study to produce grammar books that represent verb tenses with accordance to

their frequency in the naturally-occurring language by NSs.

For Students

ESL/EFL learners who want to write academic papers might appreciate knowing that the

twelve verb tenses of English do not occur with similar frequencies and therefore are not

equally relevant to their language needs. Students need to know that they do not need to use all

of the verb tenses with all their various uses all the time. ESL/EFL leaners can focus their efforts

on tenses that occur more frequently (i.e., simple past, simple present, and modals) because they

are expected to be encountered frequently. However, other verb tenses, such as present perfect

progressive tense, were not used as frequently as expected, so ESL/EFL learners do not need to dedicate as much effort to them.

Limitations

The current study sought NSs' use of verb tenses only in history academic papers. The results may not translate to NSs' use of verb tenses in their papers on another subject area, or in their spoken language.

Further Research

As far as the researcher knows, TESOL educators know relatively little about NSs' use of verb tenses; however, the findings of the current study could lead for further research studies and investigations. For example, further studies can investigate NSs' use of verb tenses in their written language on another subject area, such as science, or investigate verb tenses usage in NSs' spoken language. In addition, researchers can seek ESL/EFL use of verb tenses in their spoken/written language and compare it with that of NSs'. The results of this study can also create a baseline on which evaluations of grammar books designed for ESL/EFL learners can rely. That is, whether a grammar book writer uses the findings of corpus-based studies in representing grammar points, and whether their production rely on NSs' actual use of English.

APPENDIX A: DATA ANALYSIS

DATA ANALYSIS PAGE SAMPLE

Christianity was founded in the 1ˢᵗ century A.D. but did not become a popular religion in Western Europe until the 3ʳᵈ century. Both accounts feature a king who converted to Christianity after winning a battle they prayed they wouldn't lose. However, these two kings, Constantine and Clovis, discovered Christianity from two different sources: Constantine's father introduced him to God while Clovis's wife influenced him to convert. Although the central plot of the accounts are extremely similar, the details are very different. In The Conversion of Constantine, Constantine saw "**Conquer** by this" written in the sky at dusk. At first, he doubted whether what he saw was a sign from the heavens, but that night, Christ came to him in a dream and showed him the same sign, so he was convinced that what he saw was truly God's work and he converted. In The Conversion of Clovis, Clovis called upon God, whom he knew his wife trusted, to help him win a battle he was sure to lose. Constantine's conversion seemed like a logical tactic. He doubted God until he obtained proof of His existence (Christ in his dream). The Conversion of Clovis is not genuine; the diction is romantic in order to make the story more attractive and likeable among a greater audience. Also, there is much more detail in the grand gesture of Clovis's baptism than his actual encounter with God on the battlefield and his winning the battle.

The diction in The Conversion of Clovis is romantic and hyperbolic. The author uses words like "burning valor" to describe the soldiers after Clovis prayed to God. He also recounts Clovis looking to the heavens "humbly" when praying for a victory against the Alemanni. The author also writes that the people "believed they were already breathing the delights of paradise.

66

This word usage throughout the account implies that the author wanted to make the story of the conversion more "Hollywood," more likeable, and more interesting.

Another reason that The Conversion of Clovis isn't genuine is the amount of detail that is written in the story. After Clovis returns from battle, his wife greets him excitedly because her King and she believe in the same god now, so he gets baptized publicly at the local church. Afterwards, there is a large and extravagant parade just for his conversion. His conversion to Christianity had a great effect on his Kingdom and everyone celebrated. The amount of detail makes this account seem less genuine because it seems very extravagant for a king to convert to a religion that is not widely popular just yet. This great detail about the parade and baptism was included to make the story more interesting. In the conversion of Constantine, he returned home and appointed religious priests to his administration, therefore incorporating his newfound religion into his everyday political life and his kingdom. When Clovis returned home, there was a celebration in his honor. The account ends with a happy ending.

The two accounts of these conversions, while similar in idea, are written very differently. First, the Conversion of Constantine was written concisely. The Conversion of Clovis was drawn out and described in great detail. The whole kingdom celebrated the conversion and hosted a large and extravagant parade after witnessing his baptism into the church. Constantine had to have proof that the sign he saw on the battlefield at dusk was truly a sign from God. The only proof Clovis needed was the victory of that battle that he prayed over. The Conversion of Clovis is not a genuine account because of the descriptive diction used and the details about the grand parade and gestures made after Clovis decided to convert, giving it a magical 'Happily Ever After' ending.

APPENDIX B: THE COUNTING PROCESS ON EXCEL

SAMPLES FROM EXCEL

	A	B	C	D	E	F	G	H	I	J	K	L
1	verb tenses	p 1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11
2	simple present	15	16	15	26	55	69	74	53	20	49	18
3	present progressive	0	1	0	0	0	0	0	3	0	0	1
4	presen perfect	0	0	0	4	1	0	1	1	2	1	1
5	present perfect progressive	0	0	0	0	0	0	0	0	0	0	0
6	simpe past	39	55	34	36	17	9	24	27	48	46	43
7	past progressive	0	4	2	0	0	0	0	0	2	0	1
8	past perfect	0	0	6	6	0	0	0	1	1	5	1
9	past perfect progressive	0	0	0	0	0	0	0	0	0	0	0
10	simple future	0	1	0	0	0	5	0	1	0	2	1
11	future progressive	0	0	0	0	0	0	0	0	0	0	0
12	future perfect	0	0	0	0	0	0	0	0	0	0	0
13	future perfect progressive	0	0	0	0	0	0	0	0	0	0	0
14	modals	4	2	12	5	0	24	15	11	10	7	4
15	perfect modals	0	1	1	5	0	0	1	2	0	2	2
16	Imperatives	0	0	0	0	0	0	0	0	0	0	0

Figure 6: Counting Process on Excel-1

	p118	p119	p120	p121	p122	p123	p124	p125	p126	p127	p128	p129	p130	totals
53	34	43	63	28	11	27	39	17	66	42	16	22	57	4440
1	3	1	0	2	1	0	0	0	1	8	0	0	5	120
4	0	2	1	2	0	1	0	0	2	1	1	0	2	138
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
20	13	43	13	51	32	43	21	58	55	29	47	58	57	5002
0	0	0	0	3	0	0	0	1	1	1	2	2	0	150
2	0	4	0	0	0	0	0	2	5	0	0	0	1	183
0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
0	0	0	3	0	0	0	3	0	1	5	0	1	0	93
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	4	7	22	4	6	3	2	9	13	17	5	1	7	970
0	0	1	0	0	0	0	2	0	0	0	0	1	3	155
0	0	0	0	0	0	1	0	1	0	0	1	0	0	21

Figure 7: Counting Process on Excel-2

69

LIST OF REFERENCES

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of make in native and non-native student writing. *Applied Linguistics, 22*(2), 173-195.

Alzuhairy, U. (2016). *The frequency of the twelve verb tenses in academic papers written by native speakers* (Master's thesis). Retrieved from UCF Dissertations and Theses database (ucfl.034737242).

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes, 26*(4), 502-514.

Coxhead, A. (1998). *An academic word list*. School of Linguistics and Applied Language Studies, Victoria University of Wellington.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Folse, K. (2016). *Keys to teaching grammar to English language learners*, (2nd ed.). Ann Arbor: University of Michigan Press.

Frazier, S. (2003). A corpus analysis of would-clauses without adjacent if-clauses. *TESOL Quarterly, 37*(3), 443-466.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, *41*(2), 339-359.

Gray, B., & Biber, D. (2011). Corpus approaches to the study of discourse. In K. Hyland& B. Paltridge (Eds.), *Continuum companion to discourse analysis,* 138-154. New York, NY: Continuum International Publishing Group.

Hinkel, E. (2001). Matters of cohesion in L2 academic texts. *Applied language learning, 12*(2), 111-132.

Hinkel, E. (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language teaching research, 8*(1), 5-29.

Hunston, S., & Gill, F. (1998). Verbs observed: A corpus-driven pedagogic grammar1. *Applied Linguistics, 19*(1), 45-72.

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing, 6*(2), 183-205.

Kennedy, G. (2003). Amplifier collocations in the British national corpus: Implications for English language teaching. *TESOL Quarterly, 37*(3), 467-487. doi:10.2307/3588400

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

Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *Tesol Quarterly, 37*(4), 671-700.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes, 31*(1), 25-35.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*(1) 36-62.

Narita, M., Sato, C., & Sugiura, M. (2004). Connector usage in the English essay writing of Japanese EFL learners. *LREC 27*, 1171-1174.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223-242.

Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly, 37*(3), 419-441.

Vongpumivitch, V., Huang, J. Y., & Chang, Y. C. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes, 28*(1), 33-41.

Yoo, I. (2009). The English definite article: What ESL/EFL grammars say and what corpus findings show. *Journal of English for Academic Purposes, 8*(4), 267-278.