University of Reading

A Corpus-based Study of verb-noun collocations and verb complementation clause structures in the writing of Advanced Saudi learners of English

Thesis submitted for the degree of Doctor of Philosophy

Department of English Language and Applied Linguistics

School of Literature and Languages

Manal Abdulrahman Alangari

May 2019

Abstract

This thesis reflects the shift in the study of collocations towards lexico-grammatical patterns through a series of three corpus-based studies on academic writing. The first study adopts a phraseological approach to explore the use of adjective-noun and verb-noun lexical collocations in the academic writing of Arab learners as compared to native speakers. The comparison between two corpora, the TEEP-ArSL and the LOCNESS-A-Level, reveals that verb-noun (VN) collocations are particularly difficult for Arab learners given that more than a quarter of the VN collocations they produced are misused.

Studies 2 and 3 adopt a novel approach to the description and analysis of academic writing in the discipline of applied linguistics. These two studies focus on VN collocations by embedding them in their verb complementation clause structures, specifically noun phrase complementation. In Study 2, expert writers' use of verb complementation clause structures and VN collocations therein are examined and compared in two corpora of published research articles in the field of applied linguistics in English and Arabic, the academic English Corpus (AEC) and the academic Arabic corpus (AAC). Study 3 investigates novice writers' use of the same clause structures and VN collocations in two corpora of novice native and non-native students' university exam writing - the novice native corpus (NNC) and the novice Saudi corpus (NSC).

The analysis of the data in Studies 2 and 3 draws on Quirkian clause structures (Quirk, Greenbaum, & Leech, 1985) for the syntactic representation, and on Frame Semantics (Fillmore, 1982) for the identification of the semantic roles of the elements involved in the clause structures. The two studies explore the use of single-word and multi-word verbs in 15 clause structures, including the copular, transitive, complex copular, and ditransitive. The association between the verb and the clause structure is established using the measure of faithfulness (Römer, O'Donnell, & Ellis, 2015).

The comparison of expert academic writing in Study 2 serves as baseline data for the comparison of novice academic writing in Study 3. The results of Study 2 show that prepositional verbs are frequently used in academic English and Arabic, despite the fact that they are seldom addressed in Arabic grammar books. The results also highlight phrasal verbs as one of the prominent characteristics of current English academic writing in the field of applied linguistics.

Study 3 explores advanced learners' use of the 15 selected clause structures and the choices of verbs therein and whether there is a tendency to overuse, underuse or misuse these clause structures and VN collocations. It makes use of the results of study 2 to trace the influence of the first language, Arabic, on the use of verb complementation clause structures by advanced Saudi learners of English. The results of study 3 show that advanced Saudi learners use significantly more tokens in the copular, the transitive, and the prepositional type 1 clause structures than native speakers. However, the difference between the use of types is not significant which reveals an area of weakness in the writing of advanced Saudi learners related to the heavy reliance on a limited set of high frequency verbs, e.g. *have*, *use*, and *say*. Although native speakers used the phrasal verb clause structure more often than Saudi learners, both groups of novice writers used far fewer phrasal verbs than expert English writers; possible explanations for this underuse are investigated. The thesis concludes with a variety of pedagogical implications that could be of great benefit for language teachers and textbook and dictionary designers.

Acknowledgements

First and foremost, I would like to thank God, the Almighty, for providing me with the power and the ability to complete this work.

I also would like to express my gratitude to my supervisor, Dr. Sylvia Jaworska, for all the time and precious thoughts that she has put into this study. She provided me with constant guidance and both personal and professional support. Her most detailed comments and insightful suggestions on the drafts of this thesis were a great invaluable source of knowledge as well as encouragement. My thanks also extend to my second supervisor and PhD programme director, Dr. Jacqueline Laws, for her warm support and constructive comments. I deeply appreciate the time she spent carefully reading and revising my drafts and providing thorough feedback. Working with them has enriched me in many ways. I benefited much from their patience, welcoming positive attitude and professionalism.

I am also in debt to my country, Saudi Arabia, and Imam Abdulrahman Bin Faisal University for providing me the chance and support to pursue my studies and funding all my expenses.

My appreciation is also expressed to the module convenors, the students, and the administrators at Reading University and Imam Abdulrahman Bin Faisal University for allowing me access to the data I needed for my study. The help I received in copying and anonymising students' exam papers is acknowledged. I also must thank my colleagues who helped me in reviewing the data, especially Chris Ryder. He dedicated much of his time to revise the coding of the concordance lines and offered very organised and skilful help.

My heartfelt gratitude is to my beloved mother, Fawziah. Her endless care, love, and prayers were of great inspiration and motivation. I am also in debt to my brother, Abdullah, who was my companion and friend during my stay in the UK. Many thanks also go to my brothers, Mohammed, Adeeb, Khalid and Ahmed, my sisters, Noura, Sumaiyah and Nada, and my husband, Mohammed, for being extremely supportive and most willing to help me.

Having a baby during this journey was not easy, but my son Abdulrahman was a big source of joy and happiness. His smiles turned every hard moment into a pleasant one. His mere existence was a great motivation to work harder and be a good role model.

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Manal Alangari

Table of Contents

List of Tables

List of Figures

List of Phonetic Symbols

The transcription of Arabic examples is based on the International Phonetic Alphabet. The following table lists the Arabic sound, its usual romanisation and the IPA symbol.

Arabic sound	Usual Romanisation	IPA symbol
أ	a	ʔ
ب	b	b
ت	t	t
ث	th	θ
ج	j	dʒ
ح	h	ħ
خ	kh	χ
د	d	d
ذ	th	ð
ر	r	r
ز	z	z
س	s	s
ش	sh	ʃ
ص	s	sˤ
ض	dh	dˤ
ط	t	tˤ
ظ	dh	ðˤ
ع	a	ʕ
غ	gh	ɣ
ف	f	f
ق	q	q
ك	k	k
ل	l	l
م	m	m
ن	n	n
ه	h	h
و	w	w
ي	y	j
Means that the following syllable is stressed: /ˈʕarabiː/.		ˈ
Means that the preceding vowel is long		ː
فتحه		a
مد الالف		aː
كسره		i
مد الياء		iː
ضمه		u
مد الواو		uː

List of Abbreviations for the Glossing of Arabic Examples

1	First person
2	Second person
3	Third person
ABL	Ablative
ACC	Accusative
DEF	Definite (article)
DET	Determiner
F	Feminine
M	Masculine
NOM	Nominative
OBL	Oblique
PART	Particle
PAST	Past tense
PL	Plural
POSS	Possessive
PRES	Present tense
PROG	Progressive
SG	Singular

Chapter 1: Introduction

Collocations form an integral part of any discourse, written or spoken (Decarrico, 2001; Howarth, 1998). However, the partially restricted nature of collocations makes them very challenging for learners of English as a second language (ESL) and as a foreign language (EFL), even at advanced levels of proficiency (Al-Zahrani, 1998; Nesselhauf, 2003). ESL/EFL students' inadequate knowledge of collocations usually affects not only their comprehension of the language but also their language production (Al-Zahrani, 1998; Laufer & Waldman, 2011; Nattinger, 1988).

Research on collocations, verb-noun (VN) in particular, (e.g. Howarth, 1998; Laufer & Waldman, 2011) focuses on the lexical co-occurrence of a single-word verb with a following noun phrase, e.g. *take a risk, take a picture, make a decision, make a difference.* This lexical analysis does not take into account the particles that may follow the verb and form what is commonly known as multi-word verbs. Multi-word verbs consist of a single word verb followed by a preposition forming a prepositional verb, e.g. *depend on, look at,* or an adverbial particle forming a phrasal verb, e.g. *turn on, set up,* or both an adverbial particle and a preposition forming a phrasal-prepositional verb, e.g. *look forward to, get away with* (Quirk et al., 1985). These multi-word verbs are identified as an area of difficulty for learners of English but have been given minimum attention within the area of grammatical collocations (Alsakran, 2011; Nesselhauf, 2005).

Both lexical and grammatical collocations present a linear syntagmatic view of the use of the single- or multi-word verb and the following noun phrase. Studies based on this approach (e.g. Nesselhauf, 2005) result in lists of one-dimensional associations between two or three word classes (verb-noun or verb-particle-noun) that provide little semantic descriptions and even less syntactic variations (see Section 4.1.4 for a further discussion on collocations lists).

Verb complementation clause structures present learners with a similar difficulty to that of VN collocations, both lexical and grammatical, due to the element of 'unpredictability' (Faulhaber, Herbst, & Schmid, 2014) mostly related to the possibility of certain verbs occurring in one clause structure and not in another. For example, the verb *give* can be used in the ditransitive/double object clause structure, *I gave Mary the book*, but not the verb *explain, *I explained Mary the lesson.* The clause structure is therefore a crucial factor determining the choice of the verb and the following complementation

including the noun phrase. Expanding the analysis of collocations to include a syntactical perspective, specifically verb complementation clause structures, could lead not only to a better understanding of predictability and use of collocations; it could also contribute to a deeper paradigmatic description of language that unites both grammar and lexicon and results in a more comprehensive usage-based description of language structures (cf. Stefanowitsch & Gries, 2003).

This study takes up this perspective and presents a novel approach to the study of VN collocations, by examining which verbs co-occur, or in other words are preferred in which clause structures. The study draws on Quirkian clause structures (Quirk et al., 1985) for the syntactic analysis of the use of verbs and the following noun phrase complement, and on Frame Semantics (Fillmore, 1982) for the analysis of the semantic roles of the elements involved in the clause structures. The clause structures selected for this study are those in which the verb is complemented with a noun phrase, namely the copular, transitive, ditransitive and complex copular clause structures. The analysis of the intransitive clause structure and ergative verbs is beyond the scope of this study because in these clause structures the verb does not require complementation.

1.1 Rationale of the study

Writing in English is in demand in Saudi Arabia, especially in the context of Higher Education. English has become the medium of communication and instruction in many Saudi universities. University staff members need to write in English to communicate messages to other staff members who come from different first language backgrounds and English is the common language of communication for all of them. They also need to publish in English to advance their academic careers. It is also increasingly important for university students to write in advanced English, especially in the departments of English language and applied linguistics where students are required to use English as the means to show their knowledge and understanding of course contents.

Several studies on student English academic writing conducted in Saudi Arabia reveal that Saudi students, even at advanced levels, face problems when required to write academically (e.g. Al-Hazmi & Schofield, 2007; Al-Khairy, 2013; Alkubaidi, 2014; Grami, 2010; Javid & Umer, 2014). Most of these writing problems are attributed to two major issues:

1) Students' inappropriate lexical choices including collocations and

2) Negative discourse transfer from Arabic

When addressing the first issue, that is lexical choices, learners need to make appropriate choices when combining two or more words, that is which words co-occur with which ones. Presenting learners with lists of collocations may not lead to improvements in writings. After all, students' writing is composed of clauses and sentences (and not of lists of two or more words), in which they have to make appropriate lexical and grammatical choices. Understanding how the use of collocations depends on the clause structures could yield more fruitful insights into the actual workings of collocations in context and possibly help students formulate sentences with appropriate lexico-grammatical choices. This study, therefore, focuses on the verb, as the heart of the sentence, and its use with the following noun phrase within the clause structure.

The second issue, that is the influence of the first language on learners' comprehension and production of the target language, has been a matter of many debates in the field of Second Language Acquisition (Treffers-Daller & Sakel, 2012). Contrastive Analysis Hypothesis (CAH) advocates the notion that similarities between the native and the target language lead to a positive transfer but differences between the two languages create difficulties and lead to learners' errors (negative transfer). This notion encourages comparative studies between the first and the second language as an important step to the understanding of transfer. However, Lado's (1957) idea of transfer has been criticised because it over-predicts many errors that are not empirically observed in learners' production and fails to predict other empirically identified difficulties (Odlin, 1989: 17-18). Therefore, the behaviourist assumptions of the CAH have fallen out of favour paving the way for cognitive approaches, such as the Interlanguage Hypothesis and other Universal-Grammar based theories, such as Schwartz and Sprouse's (1996) Full Transfer/Full Access Model (Grami & Alzughaibi, 2012; Treffers-Daller & Sakel, 2012). Nevertheless, contrastive analysis has not entirely lost its appeal; it has re-appeared with the emergence of electronic corpora in the 1980s (Kramsch, 2007). Applied linguists who are interested in the contrast between language systems have made use of corpus tools to perform 'contrastive interlanguage analysis' (CIA) (Granger, 1996, 2002, 2015) between native and non-native speakers (NS/NNS) and between different groups of non-native speakers (NNS/NNS) and across different levels of proficiency. The main objective of CIA is to

enable researchers gain a better understanding of the non-native features of learners' production, not only errors but also patterns of overuse and underuse. The two terms 'overuse' and 'underuse' are used in CIA as descriptive terms to indicate that certain features are found more or less in the learner corpus (Gilquin & Paquot, 2008: 38) and point out differences that need further investigation (Lee & Chen, 2009: 285). Consequently, this understanding of learners' production can help in the design of language teaching tools and methods better tailored to learners' needs (Granger, 2002: 10; 2015: 14). At the stage of interpreting the results of CIA, classical contrastive analysis becomes very useful. A comparison between the learners' L1 and the target language facilitates the interpretation of learners' patterns of use.

This thesis adopts a CIA approach underpinned by the use of corpus tools and techniques because it best serves the research's aims, that is, 1) the identification of learners' lexical choices in relations to VN collocations embedded within the frame of verb complementation clause structures, 2) and tracing the influence of the learners' first language, Arabic. Adopting the CIA approach, this research hopes to offer useful implications for English language learning and teaching at advanced levels.

It is important to add here that in this study first language transfer is analysed based on the similarities and the differences between learners' L1 and L2 and learners' L2 performance. Such analysis may not provide all necessary and sufficient evidence for L1 transfer as it focuses on intralinguistic homogeneity which represents one effect of the three effects listed in Jarvis (2000: 253) and Jarvis and Pavlenko (2008: 35). These effects include: intralinguistic homogeneity (evidence that learners with the same L1 behave in a uniform manner when using L2); intergroup heterogeneity (evidence that learners with different L1s perform in a different way in the L2); and crosslinguistic performance congruity (the learners' use of some L2 feature parallels their use of that feature in L1) (Treffers-Daller, 2012: 54). However, to further support the evidence of negative transfer, a discussion of the results of other studies on different groups of learners with a different L1 is added when possible.

Study 1 of this thesis, as well as Study 3, attempt a NS/NNS model of CIA. The rationale for including a native speaker corpus is based on the need to identify the norms of native speakers and to demonstrate and explain how the language usage of non-native speakers differs from these norms (Cowie, 1998). Native speakers' norms may be an

unrealistic pedagogical aim, and many researchers, such as Cook (1999), argue against imposing native-speaker language, as a model, on learners of English as an international language. However, a counter argument for this criticism relies on the importance of comparison with native speakers when the aim of the study is to improve learners' proficiency because this means bringing it closer to some NS norms (Granger, 2002: 9). Furthermore, although Cook (1999) argues against foisting native-speakers' norms on learners, one cannot ignore the fact that some learners' motivation is to acquire a native-like fluency. Many studies (e.g. Jodai, Pirhadi, & Taghavi, 2014; Timmis, 2002) reveal that EFL learners have a high desire to conform to the native-speaker norms. Subtirelu (2013) reports that, even when some learners' personal preference may not necessarily be to conform to native-speakers, these learners want to do that because they want to meet other people's expectations; they do not want their non-conformity to be seen by their teachers and friends as a form of deficiency. Additionally, Ellis and Barkhuizen (2005) note that learners themselves perform cognitive comparison in the process of learning the second language because they typically have the native speakers' norms as their target. Therefore, an analysis of native speakers' norms may be regarded "psycholinguistically valid" (Ellis & Barkhuizen, 2005: 360).

In response to Granger's (2015: 17) call for a reappraisal of CIA through the inclusion of varieties of reference corpora of texts written by experts, who may or may not be native speakers, Study 2 of this thesis compares the use of verbs in the writing of expert writers in two languages, Arabic and English. This contrastive analysis of expert academic writing in English and Arabic endeavours to identify similarities and differences between the two languages in terms of the use of clause structures and verb choices. Such analysis supplies a relevant benchmark to assess and identify areas of positive or negative transfer from Arabic into English in the use of clause structures and verb-noun collocations embedded within these clause structures in academic writing by Arabic-speaking advanced learners of English. Analysing transfer based on the compatibilities and/or incompatibilities of the verb with the clause structure in learners' first language and its influence on the choices of the verbs that learners' make in the target language adds a very useful dimension to the understanding of transfer that goes beyond the consideration of semantic translations usually considered in transfer studies specifically in relation to Arab learners, (e.g. Farghal & Obiedat, 1995; Hussein, 1998).

The ultimate goal of this thesis is to suggest strategies for improving learners' lexical proficiency in academic writing in English, therefore, it is reasonable to focus on a register which is required from learners at advanced levels, and which can help them become more successful in their higher studies. Essays investigated in Study 1 are discursive/argumentative essays written for a foundation programme which addresses academic writing as a genre with common features across disciplines. However, as indicated in Bhatia (2002) and Nesi and Gardner (2012), (see Section 2.6 for further discussion), features of academic writing vary across genres and disciplines. Therefore, the selection of data for Studies 2 and 3 is narrowed down to a homogenous sample of texts from the field of applied linguistics. This data selection includes journal articles for Study 2 to represent experts writing and exam papers for Study 3 to represent novice writing.

Study 2 includes journal articles written by expert writers in English and in Arabic which represent the varieties of reference corpora mentioned in Granger (2015). Although, it is commonly recognised that published research articles are the models of academic writing, (e.g. Swales, 1990), choosing this genre as a reference against which novice writers are compared may be criticised because it is not a requirement of a university BA degree to produce work of publishable quality and it is unfair to expect students even at advanced levels to produce papers that resemble published academic writing (Nesi, Sharpling, & Ganobcsik-Williams, 2004). However, for the purpose of this study, which focuses on features related to the use of verbs in different clause structures, studying accomplished academic writing in the field of applied linguistics serves as a general model to see what kind of clause structures and therefore verb choices frequently occur in academic writing. Furthermore, comparing published academic writing in English and Arabic can act as a benchmark to better understand how the L1 of students (Arabic) might possibly influence their academic writing in English, specifically the use of clause structures, verbs and VN collocations therein.

1.2 Scope of the study

This thesis covers three studies. Study 1 represents the phraseological lexical approach to collocations. It compares the use of two types of lexical collocations, adjective-noun and verb-noun collocations in the academic essay writing of native English speakers

and Arab learners of English in order to identify the type that is more difficult for learners whose first language is Arabic.

Study 2 focuses on the description and analysis of verb complementation clause structures and VN collocations embedded within those clause structures in published journal articles written by expert writers, English and Arabic.

Study 3 investigates novice writers use of verb complementation clause structures and VN collocations and uses native and advanced Saudi students' exam papers submitted for applied linguistics modules as the source of data.

1.3 Layout of the thesis

This thesis consists of eight chapters. Chapter 2 presents a theoretical review of collocations and the dominant lexical approach used to study collocations. It moves on to explore the analytical shift that has recently been recommended to study lexis and different usage-based approaches to the study of verb complementation clause structures. Then, the representation of the target clause structures in English and Arabic is discussed in detail. Subsequently, academic writing and its features are outlined. The chapter concludes with identifying the aims and research questions of the thesis. Chapter 3 presents a description of Study 1 which led to important methodological decisions made in Study 2 and 3. The research methodology, the procedures involved in the compilation of the corpora and the steps taken to analyse the data for Studies 2 and 3 are presented in Chapter 4. Chapter 5 provides a detailed analysis and discussion of the results of Study 2, which focuses on expert academic writing, while Chapter 6 discusses the results of Study 3 on novice academic writing. In chapter 7, a general discussion is presented based on a comparison of the results of Studies 2 and 3. Chapter 8 presents a summary of the findings of the thesis, pedagogical implications and recommendations based on the findings, and concludes with outlining the limitations of this research and providing suggestions for future research.

Chapter 2: Literature Review

The present chapter explores the literature on collocations, specifically verb-noun collocations, and verb complementation clause structures. It starts with collocations in their phraseological sense at the syntagmatic level. Then it moves on to explore usage-based approaches to the study of lexico-grammatical patterns including verb complementation clause structures and the verb-noun collocations embedded within these clause structures. These usage-based approaches reflect the interdependency of lexis and grammar at the paradigmatic level. Because academic writing is the register investigated here, subsequently, academic writing and the different approaches to the study of academic writing are outlined. This chapter concludes with the aims and research questions of this thesis.

2.1 Phraseological approaches to study collocations

There are different definitions of collocations in linguistics and language teaching. The way in which collocations are defined depends very much on the researchers' approach to phraseology which is the linguistic discipline traditionally interested in word combinations, specifically collocations.

Word combinations, phraseological units, phrasal lexemes, multi-word units, and prefabricated units, or prefabs are different terms used to denote ready-made memorised combinations in written and spoken language (Cowie, 1998). According to Wray (2000), the most neutral and common term used in the literature to refer to such combinations is 'formulaic language'. However, Wray (2002) prefers the term 'formulaic sequences' because the word 'formulaic' carries with it some associations of unity, custom and habit and the word 'sequence' indicates that there is more than one distinct internal unit. Wray defines 'formulaic sequences' as: "A sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray, 2002: 9). Henceforth, the term 'formulaic sequences' is used in this study to refer to word combinations in general.

According to Cowie (1998), there are three major theoretical approaches to phraseology: the classical Russian theory, which provides systematic frameworks of descriptive categories of formulaic sequences; the anthropological approach, whose major

concern is linguo-cultural analysis of phraseology and developing conceptual frameworks for describing cultural data as represented in the meanings of formulaic sequences; and the corpus linguistic approach, which engages in the analysis of phraseology in computer-stored corpora of spoken and written language. Because the classical Russian theory and the corpus linguistic approach are more relevant to the purpose of this study than the anthropological approach, a more detailed explanation of these two approaches follows.

2.1.1 The Russian phraseological approach

The Russian phraseological theory, commonly referred to as the 'phraseological approach' (Nesselhauf, 2005: 12), has developed from the late 1940s to the 1960s. Scholars who follow this theory (Gläser, 1998; Mel'čuk, 1998) seem to agree on two primary divisions of formulaic sequences: sentence-like units and word-like units. Sentence-like units consist of word combinations that have a pragmatic function. They are identified by their role in discourse, such as: *How are you?*, *Good morning* (Howarth, 1998; Nesselhauf, 2005). Word-like units function syntactically at or below the level of the simple sentence, for example, *in the nick of time*, *a broken reed*, and *break one's journey* (Howarth, 1998; Cowie, 1998).

The Russian phraseological theory is represented in the work of Cowie (1998) and Howarth (1998). They divide word combinations into two main types: functional expressions, and composites (composite units in Howarth, 1998), corresponding to the sentence-like units and word-like units in the Russian model.

Composites, also known as collocations, can be further divided, following Benson, Benson, and Ilson (1997), into lexical and grammatical collocations depending on the word class of their elements (Howarth 1998). Grammatical collocations consist of one open class word (verb, noun, or adjective) and one closed class word (prepositions, or grammatical structures), such as *in advance* (preposition- noun), *fond of* (adjective- preposition), and *look for* (verb- preposition) (Benson et al., 1997; Howarth, 1998). In grammatical collocations, words are identified by their grammatical categories rather than by meaning association. For example, certain types of verbs can be followed by certain grammatical structures, as in: *He agreed to help me/ He avoided helping me*. The verb *agree* can only be followed by an infinitive, whereas the verb *avoid* requires a gerund after it (Al-Zahrani, 1998). Eight major types of grammatical collocations are identified in Benson et al. (1997), as G1, G2, G3, etc, see Table 2-1. The first four types are related to nouns as the principal

word in the collocation followed by a preposition (G1), a *to-infinitive* (G2), a *that*-clause (G3), and preceded by a preposition (G4). Three other types have adjectives as the main word followed by a preposition (G5), a *to-infinitive* (G6), a *that*-clause (G7). The last type (G8) is related to verbs as the main word. Under this type, nineteen English verb patterns are included, such as *SVOO, SVO to O, SVO for O…etc.*, see Table 2-2.

Table 2-1: Grammatical collocations (Benson et al. 1997)

	Collocation	**Example**
G1	Noun+ preposition	*apathy towards*
G2	Noun followed by *to + infinitive*	It was a *pleasure to do* it.
G3	Nouns followed by a *that* clause	We reached an *agreement that she would represent us in court.*
G4	Preposition + Noun	*by accident*
G5	Adjective + preposition	*angry at*
G6	Adjective followed by *to*+ infinitive	It was *necessary to work.*
G7	Adjectives followed by a *that* clause	It was *nice that he was able to come*
G8	19 verb patterns (see Table 2.2 for details)	He *sent it to him*, she *bought him a shirt.*

Table 2-2: Verb patterns for G8 (Benson et al. 1997)

Survey of Verb Patterns

In this survey the following special symbols are used: s = subject; v = verb; o = object (direct or indirect); c = complement; a = adverbial (when obligatory); v-ing = verb form in *-ing*.

Pattern Designation		Pattern	Pattern Designation		Pattern
A	=	svo *to* o (or) svoo	K	=	sv possessive v-ing
B	=	svo *to* o	L	=	sv(o) *that*-clause
C	=	svo *for* o (or) svoo	M	=	svo *to be* c
D, d	=	sv prep. o (or) svo prep. o	N	=	svoc
E	=	sv *to* inf.	O	=	svoo
F	=	sv inf.	P	=	sv(o)a
G	=	svv-ing	Q	=	sv(o) wh-word
H	=	svo *to* inf.	R	=	s(*it*)vo *to* inf. (or) s(*it*)vo *that*-clause
I	=	svo inf.	S	=	svc (adjective or noun)
J	=	svov-ing	s	=	svc (adjective)

Note that collocational types G2 and G3 (for nouns) and G6 and G7 (for adjectives) are closely related grammatically to some of the verb patterns given above.

Lexical collocations, on the other hand, consist of two open class words, such as verb-noun, or adjective- noun**.** Commonly used examples to illustrate lexical collocations

are *strong tea*, *dark night*, *blonde hair*, and *commit a crime*. The meaning association between the components of each example above is not based on grammatical rules, but on tendencies (Nofal, 2012:76). That is, *strong* and *powerful* are synonymous adjectives, but the combination **powerful tea* is not produced by native speakers, it is perceived as unacceptable (Al-Zahrani, 1998). Al-Zahrani (1998: 9) defines lexical collocations, in their simplest term, as a sequence of lexical items which may co-occur. Lexical collocations, according to Benson et al. (1997), include six types: 1- verb-noun as in *reject an appeal*; 2- adjective- noun as in *strong tea*; 3- noun-verb (naming an action) as in *blood circulate*; 4- noun 1-noun 2 as in *A colony of bees*; 5- adverb-adjective as in *deeply absorbed;* and 6- verb- adverb as in *affect deeply*, see Table 2-3

Table 2-3: Lexical collocations (Benson et al. 1997)

	Collocation	**Example**
L1	Verb + noun (verb denoting creation and/or activation)	*make an impression*
L2	Verb+ noun (verb denoting eradication and/or nullification)	*reject an appeal*
L3	Adjective + noun	*strong tea*
L4	Noun + verb (naming an action)	*blood circulate*
L5	Noun 1 + Noun 2	*a colony of bees*
L6	Adverb + adjective	*deeply absorbed*
L7	Verb- adverb	*affect deeply*

Each category, grammatical or lexical, can be divided into idiomatic and non-idiomatic. This division is not clear-cut but rather a continuum of three grades: free combinations, collocations, idioms. According to Cowie (1998) and Nesselhauf (2005), the distinction within each group of the three levels is based on two criteria: transparency and commutability. Transparency means that the elements of the combination and the combination itself have a literal or a non-literal meaning. For example, in the free combination *drink tea,* the elements are used in their literal sense, whereas, in the idioms *blow the gaff* and *under the weather,* the combination has a figurative meaning not related to the literal meanings of its elements. In collocations, the whole combination is transparent and at least one element is used in its literal sense (and at least one has a non-literal meaning). Commutability refers to the substitutability of the elements of the combination and to what degree this substitution is restricted. In free combinations, the substitutability

of the elements is restricted by their semantic properties. For example, in the combinations *read a newspaper* and *drink water*, the substitutability of *read* with *drink* is not possible because of the semantic requirement of the verb *drink* to combine with liquids. In collocations, the substitutability of the elements is not restricted by the semantic properties only. For example, the restriction of the verb *reach* to nouns like *decision, conclusion,* or *compromise*, but not *aim,* is not only related to the semantic properties of these nouns, but rather to the restricted occurrence of the verb *reach* with a limited list of nouns. Thus, the substitution of the elements of collocations is possible but is restricted. In idioms, according to Nesselhauf (2003), the substitutability of its elements is extremely limited and almost impossible.

Some researchers find these two criteria to be problematic (e.g. Nesselhauf, 2005: 25-26) because with some combinations the two criteria may not coincide. For example, in the combination *face a problem,* the word *face* is used in its figurative sense meaning 'having to deal with a particular situation' which makes it possible to classify this combination as a collocation based on the criterion of transparency. However, the verb *face* can combine with unlimited choices of objects as long as this object refers to 'some kind of difficult or unpleasant situation', such as *face a crisis, face a task, face her anger… etc.* This means that the combination *face a problem* may be classified as a free-combination based on the criterion of commutability. Therefore, to solve this problem, some researchers depend on one criterion to identify collocations, e.g. Nesselhauf (2003, 2005) who considers 'commutability' as the only criterion to define collocations, while other researchers add a third criterion which is the frequency of occurrence of the combination, discussed in the following section.

2.1.2 The corpus linguistic approach

A further highly productive approach to phraseology is the 'frequency-based approach' or 'statistically oriented approach' (Nesselhauf, 2005:12), which utilises tools developed in Corpus Linguistics. This approach extends from the work of Firth (1957; 1968) and the neo-Firthians, among whom are Halliday (1966), Sinclair (1991), McIntosh (1971) and Mitchell (1971) (cited in Nesselhauf, 2005). Firth's primary contribution is the development of the concept of collocations as an integral part of language theory. On the basis of his description of the meaning relations between words at the syntagmatic level,

collocations are no longer just an observable effect of language use, but rather a vital factor that initiates language patterns (Barnbrook, Mason, & Krishnamurthy, 2013). Sinclair (1991), a neo-Firthian, explores the concept of collocations and its relation to meaning. He defines collocations as "the occurrence of two or more words within a short space of each other in a text" (Sinclair, 1991:170). A short space, or span, is usually defined as a distance of around four words to the right and the left of the word under investigation, which is called the 'node'. Any word that occurs in the specified environment of a node is called a 'collocate'. Frequency of lexical co-occurrence is the criterion that defines collocations. The syntactic relationship between the elements does not normally play a role in deciding whether they form a collocation or not (Nesselhauf, 2005).

The neo-Firthian approach led to the burgeoning of corpus studies on collocations. These studies can be categorised into two major strands: one strand uses a corpus-driven approach and the other strand adopts a corpus-based approach. The corpus-driven approach is demonstrated in the work of Biber and his colleagues on 'lexical bundles' in general language use (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004). Biber's 'lexical bundles', a term used to refer to 'formulaic sequences', are usually not structurally complete and not idiomatic in meaning, but they serve as important discourse functions in both spoken and written registers (Biber & Barbieri, 2007). The extraction of these bundles from a corpus is based on automatic retrieval of recurrent strings of two, three or more words (Jaworska, Krummes, & Ensslin, 2015). In order to describe these bundles in discourse, Biber (2007) has developed a functional framework. Three major functional categories are identified: stance expressions, discourse organisers, and referential expressions. A series of studies followed Biber's framework, such as Hyland (2008), Juknevičienė (2009), Chen and Baker (2010) and Jaworska et al. (2015). Based entirely on frequency, this approach has the great advantage of being methodologically straightforward (O'Donnell, Römer, & Ellis, 2013). However, a high frequency of certain lexical bundles does not necessarily imply distinctive meaning and/or functions. Defining lexical bundles in terms of frequency alone results in long lists of recurrent word sequences, among which the distinction is usually based on intuition (O'Donnell et al., 2013; Wray, 2002). Due to its limitations, this approach is not adopted in this Study.

The corpus-based strand is demonstrated in the work of many research studies on learner language such as Laufer and Waldman (2011) and Altenberg and Granger (2001),

to be discussed further in Section 2.3. In corpus-based studies, types of collocations are defined and identified a priori. Then, a learner corpus is searched for instances of these types.

To sum up, collocations are defined differently based on the researchers' approach to phraseology. Cowie (1998), who represents the Russian phraseological approach, defines collocations as a type of word combination that is delimited from idioms and free combinations based on two criteria: transparency and commutability. According to the frequency-based approach, collocations are defined by Firth (1968) as the company that words keep. For Sinclair (1991), a neo-Firthian, collocations are defined as the co-occurrence of words at a certain distance and a distinction is usually made between co-occurrences based on frequency.

2.2 From collocations to lexico-grammatical patterns

Collocations, in phraseology, are identified based on three criteria, commutability, transparency (Cowie, 1998) and frequency, a criterion added by Sinclair (1991). This phraseological definition of collocations, as seen by some researchers (e.g. Stefanowitsch & Gries, 2003:210), represents syntagmatic lexical restrictions and preferences for the use of lexical items. The focus of phraseology is mainly on the lexicon; grammar is minimally considered through grammatical collocations which identify only the word class of the preceding and/or the following items (Stefanowitsch & Gries, 2003:210).

Over the last two decades, researchers have shifted the emphasis towards a more holistic view of language that emphasises the mutual relationship and interdependency between both lexicon and grammar (Stefanowitsch & Gries, 2003:210). This shift is motivated by the fact that both fields are necessary for the full understanding of the meaning and the use of a lexical item. Rather than focusing on the co-occurrence of lexical items, it is more useful for learners to be familiar with the lexical items that co-occur with a given grammatical structure (Hornby, 1954, cited in Hunston & Francis, 2000: 5).

Siepmann (2005: 430, 438) suggests loosening the definition of collocations to refer to **"**any holistic lexical, lexico-grammatical or semantic unit normally composed of two or more words which exhibits minimal recurrence within a particular discourse community." For Siepmann, the term collocations includes lexical and grammatical collocations, as well

as extended lexical units, such as the collocational relationship between the conditional clause, *if* clause, and the main clause in sentences like,

> *If one considers that the various paths do not exist except as perceived by some mind, then one immediately arrives at the conclusion that the probability of a path should be chosen proportionally to its algorithmic information* [emphasis in Siepmann, 2005:429]

The term lexico-grammar was first developed in the framework of Systematic Functional Linguistics (SFL) (Fries, 2002; Halliday, 1961; Halliday & Matthiessen, 2004). It refers to the usual lexical and grammatical environment of an item in natural texts or discourse (Gledhill, 2011: 2). Schmid (2014) defines lexico-grammatical patterns as "recurrent sequences of lexical and grammatical elements which serve an identifiable function". He classifies lexico-grammatical patterns into three groups. Group 1 consists of more fixed lexico-grammatical patterns, such as proverbs, discourse markers, verb-particle constructions (phrasal verbs, prepositional verbs, phrasal-prepositional verbs), and idioms. Group 2 includes medium fixed patterns, such as collocations and lexical bundles. And group 3 contains more variable patterns, such as valency patterns and collostructions (Schmid, 2014).

Sinclair (1991:65) is the first to demonstrate the strong association between the meanings (senses) of the lexical item and the structure in which it is used. Sinclair believes that natural language is the result of operating two principles; the idiom and the open-choice. The idiom principle states that a language user has a repertoire of semi-preconstructed phrases available for a single choice, even though they might seem analysable into segments (Sinclair, 1991:110). He observes that "the underlying unit of composition is an integrated sense-structure complex" (Sinclair, 1991:105). In other words, the process of composition is not the selection of words or structures but the selection of a unit, that has a single form and a single meaning (Hunston & Francis, 2000: 21). The idiom principle is complemented by the open-choice principle which sees language texts as a result of multiple complex choices restricted by grammaticalness. It explains language as a series of slots which can be filled with items from the lexicon. It is often referred to as a 'slot-and-filler' model (Sinclair, 1991: 109).

The concept of phrases is long discussed in the phraseological approach to collocations (e.g. Cowie, 1998; Wray, 2002) but Sinclair (1991:112), in the idiom principle,

extends the concept of phrases addressed in phraseology to include larger units of language (Hunston & Francis, 2000:21). Sinclair observes that most texts are made up of the occurrence of frequent words and if less frequent words are used, it would be in their most frequent senses (Sinclair, 1991:113). He also observes that "many uses of words and phrases show a tendency to co-occur with certain grammatical choices" (p.112), for example the phrasal verb *set out* when used to mean 'intention' is followed by a *to-infinitive* as in *Babbage set out to build a full scale working version* (p.77). Therefore, text is mostly formed by the operation of the idiom principle with occasional switching to the open-choice principle (Sinclair, 1991:113). The open-choice principle is operated in relation to lexical choices that are unexpected in their environment (p.114).

Hunston & Francis (2000: 23) demonstrate the idea of the switch between the two principles through the following extract about the hero of a novel, Arthur, who was in a spaceship and was having difficulty following the conversation of other characters:

> *Arthur blinked at the screens and felt he was missing something important. Suddenly, he realized what it was.*
>
> *'Is there any tea on this spaceship?' he asked.* [emphasis in Hunston & Francis, 2000: 23]

The most frequent sense of the verb *miss* is 'not understanding'. That is why the phrase *he was missing something* is first interpreted as 'he was not understanding'. But the subsequent question *is there any tea* leads to the reinterpretation of the phrase *he was missing something* to the less frequent sense of the verb *miss* which is 'be without'. The first interpretation resulted from the operation of the idiom principle while the second interpretation is the result of the operation of the open-choice principle.

The idiom principle refers to semi-preconstructed phrases. It raises the following question: if these semi-preconstructed phrases are not the phrases discussed in the phraseological approach and they are extended to include larger units of language, what are these units? Renouf and Sinclair (1991:128) propose the notion of 'collocational frameworks' which consist of a sequence of two words with an empty slot between them, such as 'a+?+of', 'an+?+of'. Each framework is very selective in terms of the collocates that it allows.

With the availability of corpora and corpus analysis and following from work by Sinclair (1991), usage-based theories and models such as Pattern Grammar (Hunston &

Francis, 2000), Frame Semantics (Fillmore, 2006; Fillmore, 2014, 1982; Fillmore & Kay, 1993) and Construction Grammar (Goldberg, 1995) have developed to highlight the interdependency of lexis and grammar and therefore forms and meanings in linguistic structures (Stefanowitsch & Gries, 2003:210). These meaningful grammatical structures, known as 'collocational frameworks' in Renouf and Sinclair (1991), are referred to by terms such as patterns (Hunston & Francis 2000) and constructions (Goldberg, 1995). The following sections discuss two usage-based approaches, namely Pattern Grammar and Construction Grammar whith a focus on the use of the verb and its complementation clause structures because this is the main area of interest of this thesis.

2.2.1 Pattern Grammar

Sinclair's ideas of the association of meaning and form are adopted and developed by Hunston & Francis (2000) in the notion of Pattern Grammar, in which patterns consist of a restricted set of lexical items. In order to make this approach more useful for learners, lexical items of a pattern are identified using simple word-class codes, so a pattern is presented as V n n which refers to a verb followed by two noun phrases (NPs). The verb is in upper case because it is the focus word-class of the pattern. If a preposition, an adverb, or another lexical item is part of a pattern, it is given in italics, e.g. V n *on* n. This is as far as the syntactic representation is concerned. For the semantic representation, the association between patterns and meaning is based on two pieces of evidence. First, different word senses are associated with different patterns, and, second, words used in the same pattern tend to share an aspect of meaning (Hunston & Francis 2000: 3). An example for the first piece of evidence is the verb *reflect*. When *reflect* is used in the pattern V, as in *we should give ourselves time to reflect,* its meaning has to do with thinking, whereas when it is used in the pattern V n, as in *the glass appears to reflect light naturally*, its meaning has to do with light and surfaces (Hunston & Francis, 2000: 255). As for the second piece of evidence, in the example of the pattern V *over* n, where a verb is followed by the preposition *over* and a noun phrase, many of the verbs used in this pattern share the meaning of disagreement, as in *critics argue over the niceties of translation, they disagree over the importance of the ruck duel* (Hunston & Francis, 2000: 43-44).

Pattern Grammar is taken as the starting point for a series of studies by Römer and her colleagues (Römer, O'Donnell, & Ellis, 2014; Römer et al., 2015). Their research project explores the distribution of verbs and their frequency over a large number of verb-

argument constructions (VAC) in the 100-million-word BNC corpus. The researchers identify the VAC based on Pattern Grammar, at the syntactic level. They investigate a selection of 18 'V prep n' patterns, such as *V across n* and *V about n*, of the 700 patterns identified and discussed in COBUILD Grammar Patterns 1: Verbs (Francis, Hunston, & Manning, 1996). For the semantic analyses, the researchers use WordNet which is a large lexical database of English. Then, psycholinguistic experiments are performed, in which native and non-native speakers think of the first word that comes to mind to fill the V slot in a particular VAC frame. The non-native speakers come from three different first language backgrounds, Czech, German, and Spanish. Finally, the results of the corpus analysis are compared to the results of the experiments in terms of verb selection preferences. Römer and her colleagues conclude that VACs are psychologically real in the minds of both native and non-native speakers. This conclusion is based on their finding that there is a large amount of overlap between learners' and native speakers' responses which reveal the psychological association of frames and lexical items, at least for the VACs discussed in that paper. An important implication of their study is that constructions (and phraseology in general) need to be taken more seriously in theory and practice, as their analysis indicates the inseparability of lexis and grammar.

Pattern Grammar provides a novel corpus approach to the study of language which emphasises the interdependency of lexis and grammar. However, this approach has some limitations too. It lacks the benefit of a semantic theory (Hunston, 2014). When verb patterns are concerned, less is said about the participant roles or the semantic roles of the arguments of the verb which is an important element to differentiate between clause structures which might otherwise be confused. For example, the two sentences: *She gave him a book* and *they elected her chairperson*, are presented in Pattern Grammar under one pattern (V n n) although the first sentence is an example of the ditransitive/double object clause structure (VOO), where the first NP is a recipient and the second is a theme; the latter sentence is an example of the complex copular clause structure (VOC), where the first NP is a patient and the second NP is its complement. Another example may be taken from Römer and her colleagues study of V prep n patterns (Römer et al., 2014; Römer et al., 2015). It is difficult to identify the clause structures for which the V prep n patterns correspond. The fact that in these patterns the preposition comes after a verb and is followed by a noun does not necessary mean that the verb and the preposition form a multi-word

prepositional verb followed by its object complement in a transitive clause structure. The preposition may be part of a prepositional phrase that follows an intransitive verb. For example, in the pattern V *across* n, the preposition *across* is part of the prepositional verb in the sentence *have you come across this problem?* but it is part of the prepositional phrase in the sentence *language spread across Europe.*

One of the most comprehensive attempts to add semantic elements to Pattern Grammar, according to Hunston (2014), is the FrameNet project (http://framenet.icsi.berkeley.edu, Fillmore, 2014). The FrameNet Project is a computational tool based on the theory of Frame Semantics (Fillmore, 2014). The basic idea of FrameNet is to associate lexical units with the semantic frames in which these lexical units are used. A semantic frame refers to the kinds of relations, situations or sub-events "evoked" in the minds of language users when using or encountering any of the lexical units that belonged to that frame (Fillmore, 2014). The frame elements are the things that speakers of the language would agree on the worthiness of talking about when the frame is communicatively activated, and for which the grammar provides a means of expression. For example, when a language speaker uses or encounters the verb *give,* the semantic frame of 'giving' is activated in their mind and it involves a number of frame elements including: two human participants, the donor/giver and the recipient, and an entity that is given, which is called a theme. The work of the FrameNet project on verbs initially focused on syntax only and provided a full account of the syntactic structures in which a verb is used. Later work on verbs improved the resource by including the semantic roles of the arguments associated with the verb (Fillmore, 2006).

2.2.2 Construction Grammar

In Construction Grammar (Goldberg, 1995), the focus is on constructions as a symbolic unit that represents form and meaning pairings (Croft & Cruse, 2004: 257). Following the principles of Frame Semantics (Fillmore, 1982), Goldberg argues that participant roles of a verb in a construction are derived from the event itself, for example, the participant roles of the verb *rob* are robber and victim (Goldberg, 1995:47-48). Verbs involve frame-semantic meanings; i.e., their usage must include reference to a background frame rich with world and cultural knowledge (Goldberg, 1995: 27). Constructions, as a symbolic unit, also have their own general thematic argument roles, such as agent, theme, and recipient. Verbs' participant roles must be compatible with the construction's argument

roles for the verb to be used in that construction, following the 'semantic coherence principle' (Goldberg, 1995: 50). As for the syntactic representation, Construction Grammar uses traditional functional descriptions, such as subject, object, verb, to describe the elements of a construction.

Construction Grammar has the advantage, over Pattern Grammar, of combining syntactic and semantic representations of a given construction. However, Goldberg's approach to Construction Grammar is more concerned with the analysis of the relations and links between constructions (Croft & Cruse, 2004: 272). Goldberg (1995) demonstrates that the ditransitive/ double object construction (SVOO) has a general prototypical meaning of a transfer of possession of the direct object to the indirect object (X causes Y to receive Z), as in *Joe gave Sally the ball*. However, this is not the only semantic realisation of this construction and there are a number of polysemous constructions of the ditransitive/double object construction that inherit the syntactic structure of the prototype construction but represent some semantic variations, such as *Joe permitted Bob an apple,* which has the meaning of (X enables Y to receive Z) (Croft & Cruse, 2004:274; Goldberg, 1995:75).

In literature concerned with Construction Grammar, only certain constructions in English are elaborately discussed. Some examples include, the ditransitive/ double object (Goldberg, 1995, 2006; Hilpert, 2014), the complex copular construction (SVOC) *as*-predicative (Gries, Hampe, & Schönefeld, 2005), and various complex transitive constructions (SVOA) including the resultative (Goldberg, 1995), caused-motion and the *Way* construction (Goldberg,1995; Hilpert, 2014). Based on the British component of the International Corpus of English (ICE-GB), Stefanowitsch and Gries (2003) propose a collostructional analysis of distinctive collexemes of a number of constructions such as the *into*-causative, the ditransitive, the imperative and the past tense for the purpose of evaluating and explaining the methodology. In a subsequent study, Gries and Stefanowitsch (2004), extend the use of this collostructional analysis to the comparison of pairs of semantically similar constructions, such as the ditransitive/double object construction and the *to-dative* alternation, active and passive, *will* and *be going to*. The use of collostructional analysis is extended further to a diachronic comparison of constructions by Hilpert (2006) who uses the methodology to explore the use of *shall* over three historical periods of English, as an exemplar of the analysis.

Most of these studies use Construction Grammar as an analytical framework and aim at a detailed description of a limited set of construction(s) or a comparison between syntactically synonymous constructions. To the best of the researcher's knowledge, Construction Grammar has not been used for a comprehensive description of the lexico-grammatical features of selected genres or registers and most studies are concerned with the use of English language by native speakers.

To sum up, lexico-grammatical patterns, which include larger units of language, have also attracted many researchers who are interested in collocations. However, most studies that adopt usage-based approaches to the study of these patterns focus on the analysis of the English language. Very little in said regarding learners use of these patterns (e.g. Römer et al., 2014; Römer et al., 2015). Therefore, the following Section 2.3 which explores previous studies on collocations focuses on research that adopts phraseological approach, the Russian phraseology and the frequency-based approaches, to collocations.

2.3 Collocations in Learner English

Collocations in their phraseological sense which refers to the co-occurrence of two or more lexical items have attracted many researchers who are particularly interested in learner language. This study is specifically concerned with verb-noun collocations and the syntactic clause structures of these verb-noun collocations. However, in this section studies on other types of collocations, such as adjective-noun collocations, are included to demonstrate the use of collocations in general by learners from different backgrounds.

A number of studies examine students' knowledge and use of collocations in general. These studies usually include different types of collocations and identify the most difficult type. A prominent example is a study by Biskup (1991) which focuses broadly on lexical collocations, including verb-noun collocations, and the influence of L1 on the production of collocations. Polish and German students of English are asked to provide English translations for a list of L1 collocations. The list of collocations is not provided in the published study which makes it very difficult to know the clause structures of the verb-noun collocations investigated. The results show that learners encounter great difficulties in translating verb-noun collocations from their L1 into English correctly, and that most of the errors are caused by L1 influence.

Namvar, Nor, Ibrahim, and Mustafa (2012) manually investigate Iranian learners use of English lexical and grammatical collocations in academic writing in order to detect the difficulty that learners have. The researchers code the collocations that are identified in students essays under the following categories: grammatical collocations which include verb-preposition, preposition- noun, and adjective-prepositions and lexical collocations which include verb-noun, adjective-noun, noun-noun, verb-adverb, verb-adjective, noun-adjective. Learners have produced a total of 58 grammatical collocations and 115 lexical collocations. The results indicate that learners have difficulties with both lexical and grammatical collocations in their writing with grammatical collocations being more difficult as they produced less number of them. Learners' first language seems to have a strong effect on the learners' production of collocations. Very few examples of the collocations identified in this study are provided, the provided verb-noun collocations are examples of the transitive clause structure (SVO), where the verb is followed by its object, such as *do homework, make mistake,* and *commit suicide*. The only example provided for verb-preposition collocations is the prepositional verb *rely on.*

In the context of Arabic speaking learners of English, Farghal and Obiedat (1995) and Hussein (1998) examine the use of lexical collocations by Jordanian EFL students using elicitation tests. The results of both studies show a clear deficiency in the students' knowledge of collocations. The researchers main focus is not to identify the most difficult type but rather to explore the strategies students use when they encounter unfamiliar collocations. These strategies include: the use of synonymy, paraphrasing, avoidance and/or transfer. Transfer is found to be the most commonly used strategy and it usually leads to incorrect collocations (negative transfer). Verb-noun collocations included in the tests are examples of the transitive clause structure (SVO), such as *wear her makeup,* and *have a seat.*

Two more recent studies on collocations are conducted by Shehata (2008) and Ahmed (2012) who explore Arab learners' use of different types of lexical collocations and the influence of learners' L1, Arabic, on their production using elicitation tests. Shehata (2008) investigates the effect of the environment (ESL, EFL) on the acquisition of collocations through including a total of ninety-seven Arabic-speaking participants, thirty-five students at a university in the United States, representing an ESL context, and sixty-two students from an English Department in a university in Egypt, representing an EFL

context. Two types of collocations are investigated verb-noun and adjective noun. The results show that learners in an ESL context have more advanced knowledge of collocations and less negative influence of their L1 compared to learners in EFL context. The results also indicate that, in general, learners' receptive knowledge is broader than their productive knowledge and verb-noun collocations are easier than adjective- noun collocations. Twenty verb-noun collocations are included in the test, all represent the transitive clause structure (SVO) in which the verb is followed by its object, such as *catch fire, gain weight,* except one combination, *play a role.* This combination is an example of prepositional type 3 clause structure (to be explained further in Section 2.5.3) where the verb-noun collocation is followed by a preposition. However, this syntactic difference is not identified by the researcher and the preposition *in* commonly associated with *play a role* is not included as part of the collocation.

Ahmed's study (2012) is more comprehensive as it explores six different types of lexical collocations: verb+ noun. noun+ verb, noun+ noun, adjective+ noun, verb +adverb, adverb+ adjective. The researcher uses a 60-item multiple-choice test and a 28-item translation task to analyse185 Libyan students' use of lexical collocations and the influence of the students' first language, Arabic, on their production. The findings reveal that Libyan learners have difficulty with all types of lexical collocations with adverb-adjective collocations being the most difficult and verb-noun collocations the easiest. Similar to Shehata (2008), most of the verb-noun collocations investigated in Ahmed (2012) are of the transitive clause structure (SVO), e.g. *start the car, have a bath* and *make a mistake,* the only exception is *pay attention* where the preposition *to* is not identified as part of the collocation.

Using corpus tools, Farooqui (2016) investigates the use of lexical collocations in the academic writing of non-expert, native speakers of English (NS) and non-native speakers of English (NNS), and compares it to experts' writing. The researcher focuses on the academic writing of a specific discipline, namely computer science. Two learner corpora of students' MSc dissertations, both NS and NNS, are compared with a corpus of experts' published journal articles, as a reference corpus. The results of the comparison of the 100 most frequent noun-centred and verb-centred collocations in the three corpora show that both NS and NNS overuse a large set of noun collocation (NNS (52%) and NS (78%)) and underuse a smaller set of the same type (8%) compared to expert writers. As for verb

collocations, no significant difference is found between the two non-expert corpora, NNS and NS, and the reference corpus. Farooqui (2016) uses frequency as the criterion to define lexical collocations. She defines collocations as "a node word and the word that co-occurs within the span of three words, co-occurring at least five times in total with MI-score of at least 3 and a t- score of at least 2" (p.110). The automated methodology employed results in verb-noun collocations where the syntactic and the semantic relationship between the verb and the following noun can not be clearly identified, such as *defined_VVN Section_NNW, created_VVN object_NNW,* and *ensure_VVI system_NNW.*

Another group of studies on learners' knowledge and use of collocations focuses on one type of collocations and factors that influence learners' production of this type, such as learners' level of proficiency and L1 transfer.

In his study, Liao (2010) uses elicitation techniques to investigate cross-linguistic transfer on Chinese EFL learners' knowledge of verb-noun collocations and its relationship to learners' linguistic proficiency. A 30-item multiple-choice test and a grammaticality judgment test are used to examine the extent of L1 transfer effect on three groups of participants' proficiency levels: intermediate, high intermediate, and advanced. The results reveal that learners' collocational competence correlates with their linguistic proficiency in that the more proficient learners produce more accurate collocations. Learners' competence is also significantly influenced by the effect of L1-L2 congruency, i.e. the degree of similarity between the collocations in the learners' L1 and the target L2, as similarities facilitate learning and differences restrict it. The items of the tests include verb-noun collocations of five verbs *do, have, make, take* and *get.* In all of these collocations the noun is the object of the verb, the transitive clause structure (SVO), such as *make a decision, take a job, do homework.*

Another elicitation study of verb-noun collocations is conducted by Al-Zahrani (1998) in the Saudi context. Al-Zahrani uses a 50-item cloze test to examine Saudi learners' knowledge of English verb-noun lexical collocations and the relationship between that knowledge of collocations and the students' overall language proficiency. The results reveal that there is a significant difference in learners' performance according to their academic levels, and that collocations that are non-congruent, i.e. have no Arabic equivalent, are more difficult than congruent collocations that have equivalents in Arabic. In most of the items of the cloze test the verb is followed by its object in the transitive clause structure (SVO),

such as *take a course, give a lecture, solve a problem.* However, six items demonstrate a different relationship between the verb and the following noun. For example, in the collocation *deny someone access,* the verb *deny* is a ditransitive verb followed by two objects. The collocations *pay attention, take control, attract attention* are examples of prepositional type 3 clause structure (to be explained further in Section 2.5.3). However, as in Shehata (2008) and Ahmed (2012), this syntactic difference is not identified and the prepositions associated with these verb-noun collocations, e.g. *to* in *pay attention to,* are not highlighted as part of these collocations.

Bazzaz and Samad (2011) use both elicitation techniques and learners' written production to explore Iranian learners' knowledge and use of verb-noun collocations. Data gathered from students' responses to a cloze test and manual analysis of six story writing tasks is used to quantify the relationship between learners' knowledge of verb-noun collocations and their production of this type of collocations. The results reveal a strong positive correlation as higher levels of collocational knowledge results in the use of more verb-noun collocations in written tasks. The researchers adopt Al-Zahrani's (1998) cloze test which means that the same syntactic types of verb-noun collocations are covered, mainly the transitive and prepositional type 3 clause structures.

Altenberg and Granger (2001) investigate verb-noun collocations of high frequency verbs, especially the verb *make,* in learner data as compared with native speaker data using two corpora from the International Corpus of Learner English (ICLE) (French-speaking and Swedish-speaking learners of English) and a native speaker corpus, the Louvain Corpus of Native English Essays (LOCNESS). After calculating the frequency of the verb *make* in the three corpora and testing statistical significance using the chi-square test, the results show that EFL learners, even at a high proficiency level, face difficulties when using the high frequency verb *make*. The detailed analysis which focused on the delexical use of the verb *make,* such as *make a decision,* and the causative use, as in *make sth possible* show that some of learners' difficulties are due to intralingual factors. For example, the use of the delexical verb *make* instead of another verb, as in **make a step* instead of *take a step*, or the wrong choice of the noun collocate, **make benefits*, instead of *make profits*. Many other difficulties are related to interlingual transfer, such as missing the article in the sentence **make impression* instead of *make an impression* which is the result of transfer from French. Most of the collocations investigated for the delexical use of the verb *make*

are examples of the transitive clause structure (SVO) where the verb is followed by its object. However, the causative *make* is used in the complex copular clause structure followed by an object and a complement which can be an adjective, e.g. *make something possible*, a noun, e.g. *make someone a star*, or a verb, e.g. *make someone realize something*.

Nesselhauf (2003, 2005) investigates learners' difficulties with collocations by analysing the use of verb-noun collocations in the written production of advanced German learners of English. The data used for the analysis is the German subcorpus of the aforementioned ICLE database. Dictionaries, the Oxford Advanced Learner's Dictionary (OALD, 2000) and the Collins COBUILD English Dictionary (CCED, 1995), along with a corpus analysis and native speaker judgements, are used to decide on the restriction of the collocations. Results demonstrate that, although intralingual factors have some effect on learners' production of acceptable collocations, interlingual factors seem to have a stronger influence. Non-congruent collocations, i.e. that have no equivalent in the learners' L1, pose more difficulties to learners. Nesselhauf (2005: 69) also observes that learners tend to overuse a restricted number of collocations that are also frequent in native speaker English. A possible reason of this overuse is that learners find these collocations to be safe options. If so, this could be an example of the phenomenon Hasselgren (1994) has labelled "lexical teddy bears" and which Nesselhauf (2005) refers to as "collocational teddy bears".

The verb-noun collocations identified in Nesselhauf's study are mainly examples of the transitive clause structure (SVO) (more than 75% of the produced collocations are VO), such as *give an answer, make an attempt, get a chance.* However, the list includes examples of other clause structures of which the most frequent ones are prepositional type 3 (VOPO), such as *take advantage of,* and *take sth into account,* and the double object clause structure (VOO), such as *do someone harm* (Nesselhauf, 2005: 68, 215). Nesselhauf (2005) observes that learners produce more than 30% deviant collocations for all types under investigation. However, the most frequently used three patterns (VO, VOPO, and VOO) are less susceptible to deviations, probably because they are most commonly addressed as prototypes of collocations, compared to other types that involve prepositional phrases, such as the VPO *arrive at a compromise,* that are less commonly thought of as collocations.

The use of verb-noun collocations in the writing of learners is also investigated by Laufer and Waldman (2011). A corpus of Hebrew learners of English, the Israeli Learner

Corpus of Written English (ILCoWE), at three proficiency levels have been compiled for the purpose of the analysis. Verb-noun collocations are retrieved, and their use is compared in the aforementioned LOCNESS corpus. Two comparisons are made: the frequency of collocations in the learners' corpus is compared to those of native speakers, and learners' use of the collocations is compared to other learners from other proficiency levels. The results show that Hebrew learners of English, at all three proficiency levels, produce far fewer collocations than native speakers. The mistakes made by the learners in their use of collocations are mainly due to their first language interference. L1 influence persisted even at advanced proficiency levels. The study provided no examples of the verb-noun collocations identified and no reference is made to their syntactic typology, therefore, it is not possible to evaluate the results in terms of the range of clause structures involved.

Unlike Altenberg and Granger (2001), Nesselhauf (2003) and Laufer and Waldman (2011), who use corpus tools but adopt the definition of collocations of the Russian phraseological approach that delimits collocations from free combinations and idioms based on the criteria of commutability and transparency, and like Farooqui (2016), the study of Siyanova and Schmitt (2008) uses the mutual information (MI-score) association measure as the criterion to define lexical collocations. The researchers compare NNS and NS use of adjective noun collocations using data from the Russian subcorpus of ICLE and the LOCNESS. Eight-hundred and ten adjacent adjective–noun collocations (e.g. long time) are manually extracted from the Russian subcorpus and 806 adjective–noun combinations from the native essays. The British National Corpus (BNC) of written and spoken language is consulted to determine the frequency and the MI-score of each NNS and NS collocation. The results show that Russian students produce collocations in similar frequencies to native students, and so can be considered quite successful in their production of adjective–noun collocations. Thus, the evidence obtained in this study does not support the commonly held view that L2 learners underuse native-like collocations. This study is different from other corpus-based studies because it includes the MI-score as the only measure to define collocations. Although the MI-score is considered a good measure of the strength of collocations, its use as the only criterion to decide on collocations has received criticism from some researchers (e.g. Baker, 2006; Gablasova, Brezina, & McEnery, 2017; McEnery & Hardie, 2012; Stubbs, 2001) based on the idea that the researcher's choice of the statistical test has a major effect on the outcome (McEnery & Hardie, 2012:127). MI-

score tends to favour low frequency content words (Baker, 2006: 102). It is negatively linked to frequency as it gives lower frequency combinations higher values and therefore emphasises collocations that may not be equally distributed in the corpus (Gablasova et al., 2017:164). For example, the combination *extenuating circumstances* has a very high MI-score of more than 5 for only 13 occurrences in the BNC written and spoken.

In general, the results of previous studies on learners' knowledge and use of collocations confirm that collocations form an area of difficulty for learners from different first language backgrounds, e.g. Chinese (Liao, 2010), German (Biskup, 1991; Nesselhauf, 2003), Hebrew (Laufer & Waldman, 2011), and Arab (Al-Zahrani, 1998), and that negative transfer from the learners' first language is one of the important factors behind that difficulty. In addition, this review of the previous research in collocations reveals that the main focus of verb-noun collocations has been on the SVO syntactic structure, and learner performance in relation to other frequent clause structures in English has received very little attention.

2.4 Summary of previous research

The above review starts with the definition of collocations in their phraseological sense. The discussion then moves on to lexico-grammatical patterns, as a more general concept that better reflects the association between form and meaning, and the different usage-based approaches concerned with the identification of these lexico-grammatical patterns under different labels, such as collocational frameworks, patterns and constructions. Lastly, a review of some of the studies that explore learners' knowledge and use of collocations is presented.

The review of the phraseological approaches to the study of collocations reveals some limitations of these approaches. In the Russian phraseology, collocations are defined as a type of word combination that is delimited from idioms and free combinations based on two criteria: transparency and commutability. This definition has some problems because the two criteria may correlate but do not regularly coincide, which sometimes makes it difficult to decide on the classification of a combination, such as *face a problem*, as a free-combination or a collocation. The frequency-based approach offers frequency as an automatic criterion for the identification of collocations in corpora. However, the measures of the strength of word combinations, such as t-score and MI-score, seem to have

an effect on the outcome. The t-score highlights frequent collocations while the MI-score favours low frequency collocations (Gablasova et al., 2017).

The focus of the phraseological approaches is mainly on a string of two or three lexical items, e.g. *commit a crime, blond hair,* which leads to minimal consideration of the grammatical items associated with these lexical combinations. For example, *pay attention* is defined as a verb-noun lexical collocation and the preposition *to* which is commonly used to complement the phrase is neglected. Grammatical items are given some attention under the area of grammatical collocations, such as *fond of, look at* where only the word class of the items involved is identified. The frequency-based approach attempts to extend the span of collocations through 'lexical bundles', such as *I don't know I, You know what I* (Biber et al., 2004: 384). However, this approach has been criticised for the lack of distinctive meanings for these lexical bundles.

The previously described usage-based approaches present a plausible framework that best serves the aims of this study, which is to present a comprehensive description of the use of the verb, both single- and multi-word verbs, and the following noun phrase within the frame of a larger lexico-grammatical structure in academic writing that reflects the association between form and meaning and allows the exploration of paradigmatic variations. However, examining these usage-based approaches, namely Pattern Grammar and Construction Grammar, reveals some limitations of these approaches. Pattern Grammar, as explained in Section 2.2.1, lacks the advantage of a semantic theory. It fails to highlight semantic differences between some clause structures and tends to group them together under one pattern. This limitation may be solved through the inclusion of semantic information offered by the FrameNet project. However, Pattern Grammar cannot be employed in this study for practical reasons. In Hunston and Francis (2000: 51-56), fifty verb patterns are presented under five categories; 1) verb followed by a single noun group, adjective group, or clause, e.g. Vn, V pl-n, V adj, V that; 2) verb followed by a prepositional phrase or adverb group, e.g. V prep, V adv; 3) verb followed by a noun group and another element such as another noun group, an adjective group, or a clause, e.g. Vn n, Vn adj, V n wh, 4) verb followed by a noun group and a prepositional phrase or adverb group, e.g. V n with adv, V way prep/adv; 5) verb pattern with *it*, e.g. *it* V clause, *it* V n clause. Using all of these types to investigate the use of verb-noun collocations in academic writing, is impractical to cover in a comprehensive study. Covering all these types and then presenting

recommendations for learners based on them, another aim of this study, may be overwhelming for the researcher and the learner.

Construction Grammar, on the other hand, is concerned with the detailed explanation of specific clause structures, but many clause structures that are within the area of interest of the research here, such as prepositional verbs, are not sufficiently covered which implies that there is a limitation or possibly an unsuitability of this approach if it were to be used to study a wide range of clause structures.

The survey of the literature on learners' knowledge and use of collocations also reveals a number of limitations. First, contradictory results are reported regarding the most difficult type of collocations on learners. While some researchers find that verb-noun collocations are the most difficult (e.g. Biskup, 1991), other researchers report verb-noun collocations as the easiest type of collocations (e.g. Shehata, 2008). This motivated the researcher to specifically investigate Arab learners' authentic writing to ascertain the type that is more difficult for Arab learners as a first step before conducting comprehensive analysis of verb-noun collocations and verb complementation clause structures. This step is taken in Study 1 of this thesis.

The difficulty Arab learners' encounter with collocations is identified by many researchers (e.g. Ahmed, 2012; Al-Zahrani, 1998; Hussein, 1998; Shehata, 2008) who mainly use elicitation tests as the only tool to investigate learners' knowledge and production of collocations. However, the types included in the test items are based on the researchers' predictions and choices, and students' responses to test items may not reflect what they actually know and how they use collocations in their writing. The study of Farooqui (2016) is a recent corpus-based study conducted on the writing of Arabic speaking learners of English. It has the advantage of focusing on verb-centred and noun-centred collocations. However, the results of this study are still limited to lexical collocations at the syntagmatic level. The association between grammar and meaning is not considered.

Previous research that focuses on verb-noun collocations is limited to the investigation of these collocations within the frame of one clause structure, the transitive (e.g. Liao, 2010). A few other clause structures are included, such as the prepositional type 3 (e.g. Al-Zahrani, 1998), and the complex copular in Altenberg and Granger (2001), but little is said about learner performance with these clause structures. Nesselhauf (2003; 2005) is one of the few studies that clearly identifies the syntactic patterns of verb-noun

collocations and their frequency in learners' writing. However, the list of collocations with which Nesselhauf (2005) concludes her study provides little information about the variations that may take place within the syntactic patterns identified and the frequency of these patterns. Besides, Nesselhauf mainly focuses on German learners and the type of writing explored in her studies is non-academic. This reveals that the use of different syntactic patterns of collocations and their frequency in the academic writing of Arab learners is an unexplored area.

The large body of research that describes the lexico-grammatical features of English clause structures focuses on the description of English (e.g. Hunston & Francis, 2000; Goldberg, 1995; Stefanowitsch & Gries, 2003). Very few studies have performed a comprehensive analysis of the use of verb argument clause structures in academic writing and even less has been conducted on learners' English. One of the main contributions of the current study is that it addresses this gap in the literature by evaluating learners' performance with verb-noun collocations in the context of the full range of clause structures in English.

2.5 Approach to collocations in Studies 2 and 3

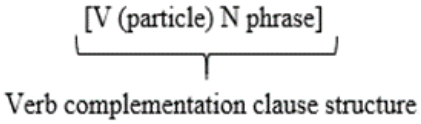
2.5.1 Definition of collocations

Usage-based approaches present a shift of focus from the linear, one-dimensional, relationships between lexical items in collocational analysis, to the focus on lexico-grammatical patterns or constructions, i.e. which lexical items could be used to fill in certain slots in the grammatical construction. This shift results in a move towards a more comprehensive usage-based description of language, particularly the use of verbs within larger syntactic structures also referred to as patterns or constructions. These lexico-grammatical patterns are formulated based on the idiom principle suggested by Sinclair (1991) and therefore present learners with an element of unpredictability which causes the same difficulty for learners as that of collocations (Faulhaber et al., 2014).

Due to the importance and the difficulty that verb complementation clause structures may pose for learners and because the researcher aims for a more comprehensive description of both single- and multi-word verb-noun collocations within the frame of verb complementation clause structures in the academic writing in the field of applied linguistics, this research follows this shift. This is also because of the fundamental role of

the verb as the heart of the sentence in determining the arguments involved in the structure, and the semantic roles of all its arguments.

Therefore, in this study, verb-noun collocations do not only refer to the lexical association between the verb and the following noun, but also the particle that may fall between the verb and the noun phrase complement is considered within the frame of a variety of clause structures that are clearly defined on the syntactic and semantic level. The concept of collocations in this study can be demonstrated as follows:



2.5.2 Quirkian Clause structures

This study employs a different approach to the study of verb-noun collocations within the frame of verb complementation clause structures. This approach systematically investigates the complementation of verbs using Quirkian clause structures (Quirk et al., 1985) as the syntactic analytical framework. Verb clause structures are explored through investigating the most frequently used verbs and the type of complementation in which these verbs most frequently occur and hence are 'preferred' by these verbs. Quirkian verb complementation clause structures present a comprehensive and concise description of language use which serves well the aims of this study.

To provide a semantic analysis of the verbs and their complementation, the researcher draws on Frame Semantics (Fillmore, 2006; Fillmore, 2014, 1982; Fillmore & Kay, 1993) in an attempt to identify the sense and the semantic frame of each verb as well as the semantic roles associated with that frame.

Because this research is interested in VN collocations, only selected clause structures are investigated. The target clause structures are mainly derived from Quirk et al (1985). Greenbaum and Quirk (1990) and Greenbaum and Nelson (2002) are also consulted. These target clause structures include the copular, transitive, ditransitive and complex copular clause structures because in these clause structures, unlike for example the intransitive, verbs are followed by a noun phrase (NP) complementation, resulting in a verb-noun combination, which is the focus of this study. The analysis includes both single word verbs and multi-word verbs used in these clause structures. Multi-word verbs are

analysed under separate clause structures, including phrasal, prepositional, and phrasal propositional variants. This special attention is justified by the fact that these structures are considered difficult for Arabic learners of English (Al-Khairy, 2013; Alsakran, 2011; Javid & Umer, 2014).

The rest of this section presents a theoretical review of the selected clause structures in English and Arabic. In this review, theoretical similarities and differences between the selected clause structures in the two languages are highlighted. It starts with a brief introduction to the verb phrase in each language then goes on to describe the representation of the clause structures in each language.

2.5.3 The Verb phrase in English

Verbs, in English, can be divided into three major categories: open class full verbs (lexical verbs) such as *leave,* closed class primary auxiliary verbs, e.g. *be, have, and do,* and modal auxiliary verbs, e.g. *will, might, etc*. Full verbs can act only as main verbs, the modal auxiliaries can act only as auxiliary verbs, and the primary auxiliary verbs can act either as main verbs or as auxiliary verbs (Quirk et al., 1985:96).

In English, verbs can be single-word verbs or multi-word verbs. Multi-word verbs are combinations of a lexical verb and a preposition, an adverbial particle or the two together; such combinations behave in certain respects as a single-word verb (Greenbaum & Nelson, 2002: 64). When a verb is followed by an adverbial particle, as in *set up, turn on,* and *hand in,* it is a phrasal verb; prepositional verbs, such as, *cope with, care for* and *look at,* are verbs followed by a preposition; when the verb is followed by an adverbial particle and a preposition, such as *look forward to,* and *get away with,* it is a phrasal-prepositional verb (Quirk et al. 1985: 1150).

2.5.3.1 Distinction based on semantic and syntactic criteria

It is important to highlight the differences between the three terms, prepositional verbs, phrasal verbs and phrasal prepositional verbs, and how they are delimited in this study from each other and from free combination. Following Quirk et al. (1985: 1163-1168) and Biber, Johansson, Leech, Conrad, and Finegan (1999: 404-428), there are a number of semantic and syntactic criteria that may be used to distinguish between the different types of multi-word verbs and set them apart from free-combinations, such as replacement by a

single-word verb, *wh*-question formation and particle movement. The semantic criterion of replacing the multi-word verb by a single word verb, according to Biber et al. (1999: 404) is more useful to distinguish between intransitive verb combinations, which are beyond the scope of this study. However, as reported by Biber et al. (1999: 404), the syntactic criterion of the *wh*-question test is more important to distinguish between free combinations with a following noun phrase and prepositional verbs, that is why the researcher depended mainly on this criterion. When the *wh*-question is formed with *where* or *when,* the combination is considered a free combination, as the noun after the preposition expresses time or location. For example, for the sentence *members are coming from Switzerland, Germany, Holland,* the question is formed as *where are the members coming from?* Therefore, the combination *coming from* is considered a free combination (Biber et al., 1999:406). When the *wh*-question is formed using *who* or *what,* the combination is classified as prepositional verb. For example, the *wh*-question for the sentence *John called on his mother* is *whom did John call on?* And for the sentence *John looked for it,* the question is formed as *what did John look for?* (Quirk et al., 1985: 1165). Therefore, the combinations *call on* and *look for* are considered prepositional verbs.

To distinguish phrasal and prepositional verbs, Quirk et al. (1985: 1167) lists a number of syntactic and phonological differences, including particle placement and movement, insertion of an adverb with prepositional verbs, and stress pattern. The syntactic criterion used in this study to delimit prepositional verbs from phrasal verbs is particle movement. Unlike prepositional verbs, the particle of the phrasal verb can be placed either before or after the object, e.g. *I put my shoes on, I put on my shoes.* In phrasal verbs, the particle cannot precede personal pronouns, if used as the noun phrase following the verb, e.g. *they called him up, *they called up him;* the relative pronoun, e.g. *the man whom they called up, * the man up whom they called;* the interrogative word in the *wh*-question, e.g. *which man did they call up, *up which man did they call?* Whereas, in prepositional verbs, the particle can precede the personal pronoun, e.g. *they called on him;* the relative pronoun, *the man on whom they called,* and the interrogative word, *on which man did they call?* In a few cases, where some phrasal verbs do not allow such alternation, e.g. *make up*, in the sentence *12 Korean stops make up the bulk of the 19 consonants,* the researcher depended on references, such as Biber et al. (1999), Quirk et al (1985), and Garnier and Schmitt (2015), to check if the verb combination is listed as an exceptional phrasal verb in these

resources. Phrasal prepositional verbs combine the characteristics of both phrasal and prepositional verbs.

2.5.3.2 Distinction based on transitivity

The term 'transitivity' is used in this study to include direct objects, indirect objects and prepositional objects. It is employed to distinguish transitive verbs from intransitive verbs that have no complementation structures. Based on transitivity, each of the multi-word verbs is divided into subtypes demonstrated in Table 2-5. The three clause structures, prepositional type 1, such as *look at*, phrasal type 2, such as *turn on/off*, and phrasal prepositional type 1, such as *put up with*, are extensions of the transitive clause structure. The phrasal type 2 clause structure is identified as such in order to set it apart from phrasal type 1, where the multi-word verb is used intransitively.

Table 2-4: Multi-word verbs in English with examples

	Phrasal Verbs	**Prepositional Verbs**	**Phrasal-prepositional Verbs**
Intransitive	Type 1: He is playing around		
Transitive	Type 2: They turned on the light	Type 1: Can you cope with the work?	Type 1: I'll get back to you.
Ditransitive		Type 2: He gave the book to Mary	Type 2: They put it down to chance.
		Type 3: Mary took care of the children	Type 3: I'll let you in on a secret.
		Type 4: Jenny thanked us for the present.	

Prepositional type 2, 3 and 4, phrasal prepositional type 2 and 3 clause structures are extensions of the ditransitive clause structure. The term 'ditransitive' is used in this research as a general term that covers single-word and multi-word verbs that involve two objects. Thus, it includes the dative and the benefactive clause structures as well as their alternations. It also includes other multi-word verbs that do not allow alternation. In the prepositional type 2 clause structure, the indirect object is introduced by a preposition, *to*

(dative) or *for* (benefactive). Two types can be identified under this clause structure; type (a), which is the alternation of the double object clause structure, and type (b), which cannot undergo alternation to the double object clause structure.

The prepositional type 3 clause structure refers to the combination in which the prepositional verb is followed by a direct object with which it forms an idiomatic fixed unit, e.g. *make fun of, catch sight of* (Greenbaum and Nelson, 2002: 66). This clause structure, referred to under other multi-word verbs in Biber et al. (1999: 427-428), includes two structures: verb+ noun phrase combination, e.g. *take care of, make fun of* and verb+ prepositional phrase, e.g. *take into account, call into question.*

When a verb is followed by an indirect object (a person who has the 'recipient' or 'affected' role) then a prepositional object, and there is no alteration to the double object clause structure, it is considered prepositional type 4a clause structure, e.g. *they told me about your success* (Greenbaum & Nelson, 2002: 66). In a similar clause structure, prepositional type 4b, the verb is followed by an object, which is not necessarily an animate, then a prepositional object, e.g. *they based the findings on fact.*

Phrasal prepositional type 2 and type 3, refer to the use of the phrasal prepositional verb with two objects. In phrasal prepositional type 2, the direct object is usually *it* followed by a prepositional object, as in *they put it down to chance.* In phrasal prepositional type 3, the indirect object is affected by the event described by the verb, e.g. *I'll let you in on a secret/ we put him up for election* (Greenbaum & Nelson, 2002: 67).

2.5.4 The Verb phrase in Arabic

Verbs, in Arabic, mainly belong to the open class full verbs (lexical verbs) and act as main verbs. A group of verbs, commonly known as 'sisters of Kana' can be used with another main verb in the same clause functioning as auxiliary verbs (Al-Aqarbeh & Al-Sarayreh, 2017: 69), as in the following example 1 where *kana* precedes the verb and its complement and gives the meaning of the past progressive tense:

(1) kaːna ʔal-ʔatˁfal-u Yu-ʃaːhid-uː-na ʔa-ˈtilfaz-a

PAST-was-3SG DEF-children-NOM PRES.PROG-watch-3PL.M.IND DEF-TV-ACC

The children were watching TV.

Modal auxiliary verbs, e.g. *will, might, etc*., which only act as auxiliary verbs in English, have no verb equivalents in Arabic. Modality in Arabic is expressed through particles, such as /sa/ which has the meaning of *will* (Wided, 2010: 23-24).

In Arabic, the verb phrase mainly consists of a single-word verb. As far as multi-word verbs are concerned, many examples of verbs followed by prepositions, such as */dalla/ (indicate) /ʕala/ (on)* and */ʔʃara/ (point to/ mention) /ila/ (to)* are common in Arabic, but they are not labelled under a specific heading by Arabic grammarians, as is the case for English (Aldahesh, 2009; 2013). Therefore, the question of whether to classify them as prepositional verbs or phrasal verbs is debatable among researchers. While Aldahesh (2009) and Lentzner (1977) believe that verb+ preposition construction in Arabic has characteristics of both prepositional and phrasal verbs, Abboud and McCarus (1968) and Heliel (1994) prefer to label them as prepositional verbs. Heliel (1994) preferred this label because when trying the syntactic test, illustrated in Section 2.5.3 above, these verb+ preposition combinations do not act as phrasal verbs but rather as prepositional verbs. Mainly, the particle, in Arabic, always precedes the noun phrase, it can never come after it. Therefore, in this study, the researcher follows Heliel (1994) and adopts the label of prepositional verbs for all instances of verb+ preposition. The researcher adopts the claim that Arabic, like other Semitic languages, has no phrasal verbs (Dagut & Laufer, 1985).

2.5.4.1 Distinction based on transitivity

Based on transitivity, prepositional verbs in Arabic can be divided into subtypes; prepositional type 1, which is an extension of the transitive clause structure; and prepositional type 2 and 4 as extensions of the ditransitive clause structure. Like in English, two types of prepositional type 2 are identified; type (a), which is the alternation of the double object clause structure, and type (b), which cannot be altered to the double object clause structure. Two types of prepositional type 4 are also identified, type (a) in which the verb is followed by an indirect object (a person who has the 'recipient' or 'affected' role) then a prepositional object, and there is no alteration to the double object clause structure; and type (b) in which the verb is followed by an object, which is not necessarily an animate, then a prepositional object.

Because of its idiomatic nature, prepositional type 3 clause structure, e.g. *make fun of, take into consideration*, is not recognised in Arabic Grammar books. However, Siinii,

Hussein and Al-ddoush (1996, cited in Aldahesh, 2009: 82-83) reported some constructions of this type, as seen in examples 2 and 3:

(2) ʔaχaða fiː ʔal-ħusban
PAST-take-3SG into DEF-account

take into account

(3) ʔaχaða bi-ˈli-ʕtibar
PAST-take-3SG into-DEF-consideration

take into consideration

Therefore, the researcher followed this approach and agreed that this clause structure exists in Arabic.

2.5.5 Verb complementation clause structures in English and Arabic

This study focuses on five main types of verb complementation clause structures, namely the copular, the transitive, the ditransitive, the complex copular, and the complex transitive. Multi-word verbs used in the verb slot in the transitive clause structure are considered under the clause structures phrasal type 2, prepositional type 1, and phrasal prepositional type 1. Multi-word verbs used in the ditransitive clause structure are identified as prepositional type 2a/b, prepositional type 3, prepositional type 4a/b, phrasal prepositional type 2, and 3. This resulted in the 15 target clause structures presented in Table 2-5.

Table 2-5: Selected clause structures (adapted from Quirk et al. 1985: 1171)

	Verb Complementation Clause Structures	**Examples**
1	**Copular** SVC (_VP NP/ ADJ) subject complement SVC 1-Adjectival 2- Nominal	 1- *The girl seemed restless.* (adjectival) 2- *She is a teacher.* (nominal)
2	**Transitive** SVO (_VP NP) 1-noun phrase (with or without passive) 2-finite clause : that-clause/ wh-clause 3- nonfinite clause: wh-infinitive, to-infinitive, -ing clause 4-to-infinitive (+s), -ing clause (+s)	 1-*Tom caught the ball/ Paul lacks confidence.* 2-*I think that we have met.* 3- *I learned how to sail/ She decided to move house/She enjoys playing squash.*

	Verb Complementation Clause Structures	Examples
		4- *They want us to help/ I hate children quarrelling.*
3	**Phrasal Type 2** SVOd (_VP AdvPart NP)	*Sam picked up the pen* *Sam picked the pen up*
4	**Prepositional Type 1** SVOp (_VP PP)	*John looked at his watch*
5	**Phrasal Prepositional Type 1** SVOp (_VP AdvPart PP)	*I look forward to your party*
6	**Complex copular** SVOC (_VP NP NP/ADJ) 1-O+ adjectival 2-O+ nominal 3-O+ to-infinitive/as 4-O+bare infinitive 5-O+ -ing clause 6-O+ -ed clause	 1- *That music drives me mad.* 2- *They named the ship 'Zeus'.* 3- *They knew him to be a spy/ He is known as a spy* 4- *I saw her leave the room.* 5- *I heard someone shouting* 6- *I got the watch repaired*
7	**Complex transitive** SVOA (_VP NP PP)	*I left the key at home*
8	**Ditransitive** **Double object** (dative, benefactive and depriving clause structures) SVOiOd (_VP NP NP) 1- 2 noun phrases 2-O+ prepositional phrase 3-O+ that-clause 4-O+ wh-clause 5-O+ wh-infinitive clause 6-O+ to-infinitive	 1- *They offered her some food.* 2- *They said something to us.* 3- *They told me that I was ill.* 4- *He asked me what time it was.* 5- *Mary showed us what to do.* 6- *I advised Mark to see a doctor.*
9	**Prepositional Type 2a** (alternating) SVOdOp (_VP NP PP)	*He lent his bike to Sam*
10	**Prepositional Type 2b** (non alternating) SVOdOp (_VP NP PP)	*He donated £10 to charity*
11	**Prepositional Type 3** SVOdOp (_VP NP PP)	*I caught sight of him*
12	**Prepositional Type 4a** (animate indirect object) SVOiOp (_VP NP PP)	*They told me about your success*

	Verb Complementation Clause Structures	Examples
13	**Prepositional Type 4b** (inanimate indirect object) SVOiOp (_VP NP PP)	*They based the findings on facts*
14	**Phrasal Prepositional Type 2** SVOdOp (_VP NP AdvPart PP)	*They put it down to chance*
15	**Phrasal Prepositional Type 3** SVOiOp (_VP NP AdvPart PP)	*They let me in on the deal*

Quirk's classification combines the SVOC and the SVOA clause structures under the same heading of complex transitive (Quirk et al. 1985: 1201-1208). However, given the semantic differences between the verb types used in these two clause structures, they are treated separately in this current research. The clause structure SVOC is referred to as the complex copular, e.g. *he considered his uncle a genius,* and SVOA as the complex transitive, e.g. *I left the keys at home.*

2.5.5.1 Copular

The copular clause structure involves a copular (or linking) verb followed by a subject complement (SVC), as in *he was hungry* / *he was a good student*, or an adverbial complement (SVA), as in *he was at school*. This post-verbal element cannot be omitted without changing the meaning or rendering the clause ungrammatical.

When the subject complement in a copular clause structure is an adjective, it may either add a current attribute or a resulting attribute to the subject of the clause structure, e.g. *The girl seemed very restless* (current), *The girl became very restless* (resulting). Current verbs, such as *be, seem, appear, keep, remain,* and *stay* are used to identify attributes that are in a continuing state of existence; they are often referred to as stative verbs. The only exception is sensory verbs, such as *feel, sound,* and *smell* which are classified as current verbs, but are used to report sensory perception (Biber et al., 1999:436). Resulting copular verbs, such as *become, get, come,* and *turn out,* are used to identify an attribute that results from a process of change (Biber et al.,1999: 436).

In Arabic, the verb */kaːna/ (copula be)* along with a group of other verbs of being, becoming, and remaining, usually referred to as 'sisters of kaːna', are used with verbless equational sentences which consist of a subject and a predicate/subject complement. As in

English, the subject complement can be a noun phrase or an adjective phrase, as in examples 4 and 5, respectively.

(4) kaːna ʔa-ˈradʒul-u muʕaˈlim-an

PAST- was- 3SG DEF-man-NOM teacher-ACC

The man was a teacher.

(5) kaːna ʔa-ˈradʒul-u sˁaːbir-an

PAST- was- 3SG DEF-man-NOM patient-ACC

The man was patient

The copular verb forms, together with the following argument, a copular or linking clause structure (VSC) similar to the English SVC (Ryding, 2005:59, 63, 634-635). As in the copular clause structure in English, the subject complement may add a current or a resulting attribute to the subject.

Hinkel (2004) points out that there are not many copular verbs in English. The most common copular verbs are *be* and *become* and they seem to prefer adjectival complements. The use of the copula *be* as the main verb is considered easier than the use of other highly semantic lexical verbs and consequently is more a characteristic of spoken language (Hinkel, 2002: 113). Hinkel (2002, 2004) observes that non-native speakers (NNS) tend to use the copula *be* in their written text more frequently than native speakers (NS).

In academic writing, the resulting verb *become* is more predominant than in any other genre (Biber et al., 1999: 443). *Become* is used to highlight the process involved in reaching a resulting state which is often related to human understanding and ideas, viewed impersonally, as in the example: *This approach becomes difficult to use and understand when a large number of entities are displayed.*

The verbs *seem* and *appear* are also common. They are used to refer to the likelihood or strong possibility and are usually employed as hedges in academic writing (Hinkel, 2004: 193; Hyland, 1998).

Saeed and Fareh (2012) investigate Arab learners' use of five verbs of senses, *look, sound, feel, taste* and *smell* using a three-part questionnaire which includes recognition, production and grammaticality judgment. The results of the study show that the students

encountered difficulties in recognizing the appropriate use of the five target verbs. Learners also have difficulty with the production of these five verbs with the verb *sound* ranked as the most difficult followed by the verbs *taste, feel, smell,* and *look* being the least difficult. The study concludes by emphasising that even advanced learners encounter difficulties in acquiring the use of high frequency verbs.

Du (2011) adopts a corpus-based approach to examine the use of copular verbs by Chinese school children learners of English as compared to native speakers. Frequency results indicate that Chinese learners significantly underuse copular verbs as compared to native speakers. Learners rely on the copula *be*; its frequency forms ninety percent of the total frequency of all copular verbs identified in the learner corpus. The researcher concludes that learners select a limited variety of copular verbs probably because they only use the verbs with which they are more familiar.

To conclude, copular verbs seem problematic for learners of English, including Arab learners who have a similar clause structure in their language. It is possible that these verbs continue to pose difficulty for learners when writing academically even at advanced levels. Since little research has been conducted to explore advanced learners' use of copular verbs in academic writing, exploring this area is important to identify areas that require improvement.

2.5.5.2 Transitive

The transitive clause structure includes, in addition to the subject, either a direct object, an indirect object or a prepositional object. The object may be a noun phrase, a finite clause such as a *that*-clause, or a nonfinite clause such as the *to*-infinitive (see Table 2-5 for examples).

It is important to note here that a small number of verbs, mainly those denoting liking or wanting, such as *bear, desire, hate, like, love, prefer, want,* and *wish,* require a subject for the non-finite clause *to*-infinitive, e.g. *They don't like the house to be left empty.* In this sentence, the NP, *the house*, is not an object of the verb *like* rather a subject of the verb *left*. This becomes evident as this NP cannot be made the subject of the passive **The house isn't liked to be left empty*. The non-finite clause *–ing* also sometimes requires a subject, e.g. *I hate the children quarrelling*. These two types of +subject clause structures require special attention because they could be confused with the complex copular (SVOC)

and the double object clause structure (SVOO): *They knew him to be a spy/ I heard someone shouting* (SVOC) or *I advised Mark to see a doctor* (SVOO) (to be explained in the following sections). The passive test can help in differentiating between these three clause structures. While the passive is not acceptable for the +subject non-finite clause object (SVO), we can easily form the passive of the clause structures SVOC and SVOO: *He was known to be a spy/ Someone was heard shouting/ Mark was advised to see a doctor*.

Type 2 phrasal verbs such as *turn on/off*, type 1 prepositional verbs such as *look at*, and type 1 phrasal-prepositional verbs such as *put up with* are the types of multi-word verbs that can fill in the verb slot of single-word verbs (Greenbaum & Quirk, 1990: 344-345).

In Arabic, the transitive clause structure involves a transitive verb that takes a direct object. The subject and the object of the verb usually come after the verb. They may be a noun, a pronoun, or an inflectional suffix attached to the verb pronoun. The verb and its complements, the subject and the object, form the clause structure VSO which is an equivalent to the English SVO. In example 6, the verb */ba:ʕ-a/* (sold) is followed by the subject /*l-muhandis-u/* (the engineer) and the object */s-sayya:rat-a*/ (the car) . In example 7, the subject of the verb takes the form of an inflectional pronoun /*tu*/(I) attached to the verb followed by another attached pronoun "hu" (it) which refers to the object that has been written. The subject also could be covert and understood from the context as in example 8, where the pronoun (he), which refers to the subject, is covert in Arabic and there is an attached pronoun "hu" (it) which refers to the object that has been found by him.

(6) ba:ʕ-a ʔal-muhandis-u ʔa-ˈsaja:rat-a

PAST-sold-3SG DEF-engineer-NOM DEF-car-Acc

The engineer sold the car. (Salih, 1985: 52)

(7) katab-tu-hu

PAST-write-1SG-it-3SG

I wrote it.

(8) wadʒada-hu

PAST-find-3SG-it-3SG

He found it.

The object could also be a non-finite clause, expressed in Arabic using the particle */ʔan/* followed by the verb in the subjunctive mood (see example 9). This clause could be

altered to a verbal noun as in example 10. The object could also be a finite clause starting with the particle relative pronoun */ʔaˈna/*, as in example 11.

(9) ʔaħtadʒ-u ʔan tusa:ʔida-ni

PRES-need-1SG to PRES-help-2SG-me

I need you to help me.

(10) ʔaħtadʒ-u musa:ʕadatak-a

PRES-need-1SG help-2M- POSS.ACC

I need your help

(11) taʕni ʔaˈna-hum sˤaːbiruːna

PRES-mean-1SG that- 3PL-they patient-PL.NOM

It means that they are patient

As discussed earlier in Section 2.5.4, the researcher adopts the claim that phrasal verbs do not exist in Arabic. Therefore, phrasal verbs type 2 and phrasal prepositional type 1 clause structure are not identified in Arabic. However, prepositional verbs exist in Arabic and are used as multi-word verbs that fill the verb slot in the transitive clause structure resulting in prepositional type 1 clause structure.

According to Biber et al. (1999: 382-3), the majority of commonly used verbs occur in the transitive clause structure, and they are from all semantic classes. Biber et al. (1999: 360-1) identifies seven major classes for verbs including: activity verbs, communication verbs, mental verbs, causative verbs, verbs of simple occurrence, verbs of existence or relationship, and aspectual verbs. Hinkel (2004: 178) suggests that only five semantic classes are particularly relevant to academic writing and useful for the teaching of English to second language learners (L2). These semantic classes are activity verbs (*make, use, show*), reporting verbs (*explain, discuss, argue*), mental verbs (*know, consider, see*), linking verbs (*appear, become, remain*), and logical-semantic relationship verbs (*change, follow, cause*).

Given the fact that this study focuses on academic writing and has pedagogical aims, Hinkel's (2004) classification of verb semantic classes is adopted, see section 4.1.4.

The semantic class of linking verbs mainly covers copular linking verbs; that is why it is excluded while classifying the verbs in the transitive clause structure.

One of the most comprehensive studies on lexical verbs in academic writing has been conducted by Granger and Paquot (2009). The study provides a detailed description of learners' use of lexical verbs compared to the use of expert and novice native writers. The learner data comes from the second edition of the International Corpus of Learner English (ICLE) and includes argumentative essays written by high-intermediate to advanced EFL university students from 16 different mother tongue backgrounds, such as Bulgarian, Chinese, Czech, Dutch, Finnish, French, and German. The study covers lexical verbs from different semantic classes. The findings indicate that learners significantly underuse academic verbs that express a logical-semantic relationship, such as *relate* and *include*, and depend more on less academic verbs, such as *think* and *like* which are more associated with conversational informal speech. The researchers also observe through their analysis of the lexico-grammatical patterns of the verbs *conclude* and *argue* that learners use a very limited range of patterns compared to the rich patterning found in experts' writing.

Given the importance of reporting verbs in academic writing (Hyland, 1998), this semantic class of verbs has received considerable attention in research. Zhang (2008) and Manan and Noor (2014) both explore the use of reporting verbs in Master theses. Zhang compares two corpora of 40 theses, each, written by Chinese learners of English and English native speakers and generally concludes that English writers use more reporting statements than Chinese writers. Manan & Noor study reporting verbs in 6 theses written by Malaysian students. The analysis show that learners depend on verbs that simply report the findings of previous researchers. Verbs used to express critical thinking, synthesising, criticising are less frequently used.

Using a corpus-based approach, Bloch (2010) investigates the use of reporting verbs in academic papers. The researcher compares three corpora, two are composed of articles from the journal "Science", and the third corpus consists of students' essays, unfortunately the L1 background of the students is not mentioned, in order to compare the learners' use of reporting verbs to that of expert writers. Bloch (2010: 221) observes that learners of English experience difficulty in choosing the reporting verbs that can "satisfy both the syntactic requirements of their sentences and, perhaps more importantly, to express their

attitudes towards the claims". Learners depend on a limited set of reporting verbs which they use repeatedly and choose them randomly without considering the impact and the function of the reporting verb.

As far as prepositional verbs are concerned, Zareva (2016) explores native speakers' use of prepositional, phrasal and phrasal-prepositional verbs in oral presentations as part of a university course requirement. The results of the study reveal that prepositional verbs are twice as frequent as phrasal verbs, followed by the relatively infrequent use of phrasal prepositional verbs.

Learners' use of prepositional verbs has been investigated in a number of studies. For example, Wilcoxon (2014) explores Spanish learners' use of prepositional verbs in English academic writing produced in the context of university courses and compares the frequencies of these verbs across proficiency levels. The prepositional verbs used by learners are classified based on the categorisation labels provided in *Longman Grammar of Spoken and Written English* (Biber et al., 1999), namely academic (ACAD), fiction (FICT), news (NEWS), and conversation (CONV). The results show that Spanish learners use prepositional verbs with high accuracy and they use both academic and conversational prepositional verbs. The researcher finds it encouraging that students' use of academic prepositional verbs increases as their proficiency level improves.

Al-Amro (2006) and Alsakran (2011) investigate Arab learners' use of verb-preposition collocations using a 16-item gap-filling test. The results of the two studies on Arab learners report on learners' low performance in the production of verb-preposition collocations. However, it is important to mention that the researchers seem to be using the term preposition in verb-preposition collocations to refer to any particle after the verb without discriminating between prepositional verbs and phrasal verbs. This could be attributed to their use of the Benson et al. (1997) classification of lexical and grammatical collocations, see Tables 2-1 and 2-2, where no clear distinction is made between prepositional verbs and phrasal verbs.

The importance of reporting verbs, as one of the semantic classes of verbs used in the transitive clause structure, and prepositional verbs is clearly identified in the literature. However, other semantic classes of verbs in the transitive clause structure, such as mental, activity and logical semantic relationship verbs, are given less attention. Even less is said about Arab learners' use of reporting verbs and/or other semantic classes in the transitive

clause structure. Prepositional verbs are identified as an area of difficulty. Nevertheless, further research which clearly defines prepositional verbs, delimits them from phrasal verbs and examines their use in academic writing is needed.

2.5.5.3 Complex copular

In the complex copular clause structure, there are two elements after the verb that have an object-predicate relationship. In the sentence: *She considered her mother a sensible woman, her mother* is the object of the verb *consider*, but the subject of the copular clause *her mother is a sensible woman*. The verb is followed by a direct object and an object complement, which can be an adjective or a noun phrase, e.g. *The secretary left all the letters unopened* (adj). *I have often wished myself a millionaire* (noun). The object complement could also be a clause: *to*-infinitive, bare infinitive, -*ing* clause, -*ed* clause (see table 2-5 for examples). The most common verbs in this clause structure are listed by Greenbaum & Quirk (1990: 350) and include *appoint, believe, call, choose, consider, declare, elect, find, get, like, make, name, prefer, think, want*.

In Arabic, verbs that have the sense of perception or transformation are known in Arabic grammar books as 'the sister of Dhanna' /zˤɑnɑ/. They take two objects, but these objects have an object-predicate relationship. These two objects form a kind of a finite clause where a relative pronoun is omitted, as in the SVOC clause structure in English.
The sisters of Dhanna /zˤɑnɑ/ fall into two classes: 1- verbs that have the sense of perception, judging, and considering, (see example 12), and 2- verbs that have the sense of transformation, making, creating, naming, and appointing as in Example 13 (Ryding, 2011: 286).

(12) ðˤanant-u ʔabaːk-a muʕaˈlim-an

PAST-consider-1SG father-POSS-2SG-ACC teacher-ACC

I considered your father a teacher.

(13) ʤaʕalt-u Ali mudiːr-an li-ʃ-ˈʃarika-ti

PAST-consider-1SG-I Ali-ACC manager-ACC of- DEF-company-OBL

I considered Ali the manager of the company.

2.5.5.4 Complex transitive

When the direct object is followed by an adverbial complement which is usually a prepositional phrase of direction or metaphorical extension of the notion of direction, specifically with "put" or "accompany" verbs, such clause structure is identified as the complex transitive clause structure. e.g. *I slipped the key into the lock, May I see you to your seat?* (Greenbaum & Quirk, 1990: 350-351). Space position particles also occur in this pattern, e.g. *Always keep your eyes on the road when driving* (Greenbaum & Quirk, 1990: 351). The complex transitive clause structure (SVOA), where a verb is followed by an object and an adverbial complement, is possible in Arabic (Lentzner, 1977: 200), as in example 14.

(14) wadˤaʕat-u ʔal-kitaːb-a ʕala ʔal-maktab-i

PAST-put- 3SG DEF-book-NOM on DEF-desk-OBL

I put the book on the desk

A number of studies are conducted on the complex copular and the complex transitive clause structures within the framework of Construction Grammar (e.g. Goldberg, 1995; Gries et al., 2005; Hilpert, 2014; Stefanowitsch & Gries, 2003). These studies are mainly concerned with describing the native speakers' use of this clause structure. To the best of the researcher's knowledge, learners' use of the complex copular and the complex transitive clause structures is yet to be explored.

2.5.5.5 Ditransitive

The ditransitive is used in this study, following Quirk et al. (1985), as a general term that covers all the clause structures in which the verb is followed by two objects including multi-word verbs. In the ditransitive/double object clause structure, the verb is followed by two object noun phrases: an indirect object, which is normally animate and positioned first, and a direct object, which is normally inanimate, e.g. *He gave the girl a doll* (Greenbaum & Quirk, 1990: 353). This clause structure can alternate to form the prepositional type 2a clause structure which includes a prepositional object as the second object, usually with *to* or *for*, so the structure becomes *he gave a doll to the girl*. Examples of these verbs that allow both structures are: *bring, deny, hand, lend, offer, owe, promise, read, send, show, teach* and *throw*, which take *to* and *find, make, order, save,* and *spare*, which take *for*.

Sometimes the prepositional clause structure cannot alternate to the double object clause structure, such as *He donated £10 to charity*, this type is classified as prepositional type 2b

Prepositional type 3 and type 4a and b, and type 2 and 3 phrasal-prepositional verbs are also included in this category (Table 3-1). Greenbaum and Nelson (2002: 66) label the combination in which the prepositional verb is followed by a direct object with which it forms an idiomatic unit, and allows no alteration, as type 3 prepositional verbs, e.g. *make fun of, catch sight of.*

When a verb is followed by an indirect object (a person who has the 'recipient' or 'affected' role) then a prepositional object, and there is no alteration to the double object clause structure, it is considered a type 4a prepositional verb, e.g. *they told me about your success* (Greenbaum & Nelson, 2002: 66). In a similar clause structure, prepositional type 4b, the verb is followed by an object, which is not necessarily animate, then a prepositional object, e.g. *They based the findings on fact*.

In ditransitive phrasal-prepositional verbs type 2, the first object is an unaffected theme, often *it*, which forms part of a fixed expression, e.g: *Don't take it out on me*. Whereas in type 3, the object is affected by the verb, e.g. *they let me in on the deal.*

In Arabic, a number of verbs which have the sense of 'giving' take two objects (Ryding, 2011: 286). These verbs along with their objects form a clause structure similar to the ditransitive/double object clause structure in English (SVOiOd). The indirect object is the beneficiary of the action, as in example 15. When the direct object comes before the indirect object, the indirect object is preceded by a preposition *to*, or *for* as in example 16, forming a clause structure similar to prepositional type 2a in English (SVOdOp).

(15) ʔaʕtˁa ʔal -malik-u ʔa-ˈrajul-a kita:b-an
PAST-give-3SG DEF-king-NoM DEF-man-ACC book-ACC
The king gave the man a book. (Danks, 2011: 106)

(16) ʔaʕtˁa ʔal-malik-u kita:b-an li-r-ˈrajul-i
PAST-give-3SG DEF-king-NOM book-ACC to-DEF-man-OBL
The king gave a book to the man.

The same prepositional clause structures, prepositional type 2b, prepositional type 3 and prepositional type 4a and b, explained above, are found in Arabic. Prepositional type 3 are not very frequent due to their idiomatic nature but this clause structure is identified in

Arabic following Siinii, Hussein and Al-ddoush (1996, cited in Aldahesh, 2009: 82-83), as explained in section 2.5.4.

Many studies have focused on the double object clause structure and its prepositional type 2a alternation exploring its use in the English language by native speakers within the framework of Construction Grammar, e.g. (Beermann, 2001; Goldberg, 1995; Goldberg, 2002; Gries & Stefanowitsch, 2004; Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989; Stefanowitsch & Gries, 2003). Little research has been conducted, however, on learners' use of these two clause structures and on the other ditransitive clause structures with prepositional verbs.

Phrasal verbs are discussed separately in the following section because they do not exist in Arabic. These verbs fill in the verb slot in the phrasal type 2 and phrasal prepositional type 1,2 and 3 clause structures.

2.5.5.6 Phrasal verbs

This section zooms in on the use of verb-noun collocations in the following structures: phrasal verbs type 2, phrasal prepositional type 1,2, and 3, and prepositional type 3 clause structures in academic English writing. This group of clause structures is chosen for a detailed analysis for a number of reasons. First, these clause structures do not exist in Arabic, therefore, they may cause difficulty for Arab learners. Research concerned with the use of phrasal verbs by learners showed that they prove to be error-prone for learners of English in general (Paquot & Granger, 2012: 133). An extensive analysis of the use of phrasal verbs in these clause structures, their meanings, and the semantic roles involved may provide useful insights for learners. Another reason is related to the importance and frequency of phrasal verbs in the English language. Based on a corpus search of phrasal verbs in the British National Corpus (BNC), Gardner and Davies (2007) report that learners may encounter an average of one phrasal verb in every 150 words of English.

Previous literature on English phrasal verbs has focused mainly on identifying and examining the most frequent verbs in general English usage and their senses. For example, Gardner and Davies (2007) investigate the most frequent phrasal verbs and their senses in the BNC. The results of their study included a list of 100 phrasal verbs, each with an average of 5.6 meaning senses. For example, *go on* has 5 senses, *carry out* (2), *set up* (12),

pick up (16), and *point out* (3). Liu (2011) uses the Corpus of Contemporary American English (COCA) to re-examine the use of the 100-phrasal verb list proposed by Gardner and Davies (2007), with an addition of 50 more phrasal verbs. The study endeavours to compare the use of English phrasal verbs a cross English varieties and registers. Liu (2011) concludes that the most frequent phrasal verbs in British and American English are similar. The cross-register analysis reveal that phrasal verbs are more common in fiction and conversation than in academic writing. Garnier and Schmitt (2015) develop a pedagogical list of English phrasal verbs (the PHaVE list). This list includes the 150 most frequent verbs identified in Liu (2011) and their key senses along with the percentage of occurrence of each of the key senses, definitions and examples. The researchers find that only two senses on average are enough to cover three-quarters of the occurrences of each phrasal verb.

All of these studies have tackled one side of the study of phrasal verbs which is to list the most frequent items and identify their most important senses. Little is said about the argument structures of these verbs and the semantic roles of the elements involved in these structures. Additionally, these studies are based on the BNC and COCA, therefore, the results and the senses reported in previous research are mainly referring to general usage of phrasal verbs rather than academic usage. This gap identified in the literature is the third reason behind the focus on phrasal verbs in academic writing in this study.

Learners' use of phrasal verbs has also attracted attention in research. Many studies use multiple-choice tests to assess learners' knowledge and production. Alshayban (2018) and Aldukhayel (2014) collected data from undergraduate and graduate Saudi learners of English. Alshayban (2018) finds that learners are reluctant to use phrasal verbs, especially when communicating with native speakers, because of their fear of misusing the phrasal verbs and consequently being misunderstood. Aldukhayel (2014) observes that students' exposure to the English language influences their production of phrasal verbs; longer exposure results in better production.

Mahmoud (2015) manually checked 84 essays written by Arabic university students. The researcher noted an under-representation of phrasal verbs in University students' writing that can be mainly attributed to passive learning. Abdul Rahman and Abid (2014) investigated Omani, first and fourth year, university students' use of phrasal verbs using two recognition tasks and one free-writing task. The results of the study showed that phrasal verbs were rare or non-existent in Omani students' writing. El-Dakhs (2016)

involved 407 Egyptian university undergraduates in her study and asked them to complete a paraphrase task, two gap-filling tasks and a survey. The results indicate that learners tend to avoid phrasal verbs in production. This avoidance is caused by a variety of factors including the idiomaticity of some phrasal verbs and cross-linguistic differences.

2.5.6 Summary related to Quirkian clause structures

The above theoretical review of Quirkian clause structures and their representation in English and Arabic reveals many similarities between the two languages. The five selected main clause structures, namely the copular, transitive, complex copular, complex transitive and ditransitive, similarly exist in the two languages. Phrasal verbs, which can fill in the verb slot in the transitive and the ditransitive clause structures in English, form an example of multi-word verbs that do not exist in Arabic.

The survey of the literature on each clause structure, presented under section 2.5.5, reveals that Arab learners' use of these clause structures in English is still an under-researched area. Few studies, sometimes none as in the case of the complex copular, are found for every clause structure. Furthermore, to the best of the researcher's knowledge, none of the researchers have attempted a large-scale study that covers learners' use of clause structures and verb choices therein. It is also generally noted that most studies on Arab and Saudi students use multiple-choice tests (e.g. Alsakran, 2011; Saeed and Fareh, 2012; El-dakhs, 2015) as the empirical measure of learners' production. Very few studies have to date examined Arabic learners authentic written production (e.g. Mahmoud, 2015). Contrastive and comparative analyses of clause structures and the choice of verbs within these clause structures are still outstanding.

Before the research aims and research questions are defined, the following section discusses features of academic writing because this register is the focus of this study.

2.6 Academic writing

Academic writing is a broad notion that covers different types of writing including journal articles, theses, examination papers, essays and other written work (Bhatia, 2002: 28). In applied linguistics literature, academic writing has been regarded a unified entity which has 'a common core' across genres and disciplines and consequently courses of English for academic purposes (EAP) and textbooks have been designed to describe the common features of academic writing (Bhatia, 2002:25). However, even though many of

those textbooks are commercially successful, it is difficult to assert that they are based on principled research findings rather than the experience and/or perceptions of the author (Bhatia, 2002:25). They are also very general. Rarely differences in academic writing across disciplines are concerned though we know that scientists write very differently from social scientists or scholars based in arts and humanities. Recent research on genre analysis reveals that academic writing varies considerably across disciplines and each discipline favours particular genres. For example, Nesi and Gardner (2012) distinguish between thirteen different assignment types that also vary according to disciplinary norms and requirements. Nesi and Gardner use the term 'genre family' to include the variations of a genre across different disciplines. For example, the Critique genre family includes book reviews from the History discipline, evaluation of research methods from the Sociology discipline as well as evaluation of production techniques from the Engineering discipline. This variation makes it necessary for applied linguists and educators to shift from the traditional focus on common core analysis, to a more broad and flexible view of language use which appreciates the differences between genres and disciplines that fall under the umbrella of academic writing, or academic discourse, in general. Such consideration may lead to a more effective pedagogical material (Bhatia, 2002: 30).

A distinction must be made here between register and genre. Biber and Conrad (2009: 2) distinguish between the two terms based on the perspective of the research. Research that explores the conventional structure of a complete text, for example the conventional way of writing thesis abstracts, is an example of genre analysis; research that explores specific linguistic features of a text, for example the use of verbs or pronouns is concerned with a register. Considering this distinction, this research is interested in academic writing as a register, as it endeavours to explore the lexico-grammatical features of academic writing.

Three main approaches have been adopted to study academic writing: Genre analysis which focuses on the logical appropriate organisation of ideas in a text, the 'generic elements' or 'moves' of that text (Charles, Hunston, & Pecorari, 2009:2), Register analysis approach which focuses on the syntactic and lexical features of academic texts, and contrastive analysis which focuses on the comparison of learners writing (L2) to the writing of natives or experts (L1) (Charles et al., 2009:1-3; Hinkel, 2002:15-20).

Genre analysis has been pioneered by the work of Swales (1981; 1990; 2004) who is the first to introduce genre analysis to academic writing through his analysis of two genres, research articles and thesis introduction. Following Swales, Bhatia (1992, 1993) investigates articles and other types of discourse in various academic disciplines, such as psychology, sociology, and law.

Register analysis began to flourish with the introduction of corpus studies and the use of corpus tools in language analysis. Corpus tools facilitate the examination of large amounts of authentic data from different texts and provide frequency and distributional information about some features of the language (Charles et al., 2009). Based on corpus-based studies, many grammar books and learner dictionaries which include description of lexis and collocations have become available. For example, Sinclair and colleagues started the Collins COBUILD English project which led to the publication of the *Collins Cobuild English Grammar* (1990) and Biber and colleagues published the *Longman Grammar of Spoken and written English* (LGSWE) (Biber et al., 1999). The 'big four' dictionaries (De Schryver, 2012), also became available by the year 1995, as new editions of the Oxford advanced learner's dictionary, the Longman dictionary of contemporary English, the COBUILD English dictionary were published and the Cambridge international dictionary of English was first launched. Learners' use of dictionaries and its effectiveness is to be discussed further in chapter 8.

Following register analysis, Biber (1988) used a multi-feature/multi-dimensional (MF/MD) approach to explore the frequencies of specific linguistic features used in written and spoken texts and identifies their co-occurrence patterns. Biber's dimensions refer to "bundles of linguistic features that co-occur in texts because they work together to mark some common underlying function" (Biber, 1988: 55). That results of Biber's analysis of a corpus of a variety of genres, including academic prose, official documents, broadcast, and conversation, and subgenres, such as natural science, medicine, mathematics, social and behavioral sciences, political science, humanities, and technology as subgenres for academic prose, reveals that sub-genre differences are responsible for the variation observed between the major genres, especially for the case of academic writing. The seven academic sub-genres are different from each other with respect to each of the dimensions investigated (Biber, 1988: 198). Biber's conclusions further support Bhatia's (2002) and

Nesi and Gardner's (2012) observations against 'common core' analysis of academic discourse.

It is important and specifically relevant to the topic of this research to mention, here, Biber's later work on lexical bundles (Biber & Barbieri, 2007; Biber et al., 2004) which emphasises the importance of multi-word sequences in academic university written and spoken registers. Biber's lexical bundles have been taken up by Hyland (2008) to explore 4-word bundles in three genres: research articles, doctoral dissertations and Master's theses, in four disciplines, Biology, Electrical engineering, Applied linguistics, and business studies. Hyland's analysis shows that bundles form essential building blocks of academic discourse and they differ across disciplines which means that they can be used to characterize disciplinary discourse.

A large body of corpus studies focuses on contrastive analysis of academic writing. These studies are centred on learner corpora and explore the grammatical or lexical features overused or underused in the writing of non-native speakers (NNS). NNS corpora are usually compared with parallel native-speaker (NS) corpora, or sometimes with expert corpora. This approach has been pioneered by the work of Granger (Granger, 1998b, 2003), that led to the compilation of the International Corpus of Learner English (ICLE) (Granger, Dagneaux, Meunier, & Paquot, 2009). This large corpus of essays written by students from different L1 backgrounds, Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana, and Turkish has been used in many contrastive corpus studies. Granger's book (1998b), *Learner English on Computer*, includes several papers that used the ICLE to represent a learner corpus and compares it to the Louvain Corpus of Native English Essays (LOCNESS). For example, Ringbom (1998) investigates the frequency of high frequency verbs. Granger and Rayson (1998) study the frequency of different word-classes in the ICLE and LOCNESS corpora. Biber and Reppen (1998) compare native and learner use of complement clauses. These studies highlight some of the characteristics of learner English academic writing and the factors that influence learners' production including learners' L1 as well as spoken registers.

A number of studies following the contrastive analysis approach focus on comparing the use of collocations, phraseology and formulaic sequences by NNS and NS.

Studies reviewed in section 2.3., such as Altenberg and Granger (2001), Nesselhauf (2003) Laufer and Waldman (2011) and Siyanova and Schmitt (2008) are some examples.

Hinkel (2002) explores 68 linguistic and rhetorical features in the writing of NNS compared to NS. NNS are speakers of six languages: Chinese, Japanese, Korean, Vietnamese, Indonesian, and Arabic. The features include linguistic features, such as semantic and lexical classes of nouns, personal pronouns, verb tenses, modal verbs, and rhetorical features, such as hedges, and demonstrative pronouns.

Studies on learners of English reveal some general features of NNS academic writing. For example, learners tend to use a smaller range of vocabulary items and certain general vocabulary, such as *people* and *things* (Hinkel, 2002; Ringbom, 1998). Learners use features which are typically associated with spoken English rather than academic written English, such as the use of *be* as a main verb (Hinkel, 2002; 2003), the overuse of first and second person pronouns (Granger & Rayson, 1998), the overuse of infinitives (Granger and Rayson 1998), and the underuse of nouns (Granger and Rayson 1998). In their study of complementation clauses, *that*-clauses, *to*-clauses, *ing*-clauses and *wh*-clauses, following verbs like *think, say, know, show,* and *hope,* Biber and Reppen (1998) found that learners overuse *that*-clauses and *to*-clauses compared to native speakers. They concluded that there is a similarity between the patterns used in learners' academic writing and those used in native conversation and fiction, but not academic writing.

Some features of the writing of Arabic-speaking learners of English are explored by Ostler (1987) and Sa'adeddin (1989), reviewed in Hinkel (2002:38), but these two studies mainly focused on rhetorical features which are not particularly relevant to the research at hand. Hinkel (2002, 2003) included a group of Arabic-speaking learners of English in her study, and the results reveal some linguistic features of Arabic-speaking learners of English. Hinkel (2003) observes that NNS employ simple syntactic and lexical constructions, such as *be*-copular as the main verb, predicative adjectives, vague nouns, and public, private and expecting/tentative verbs, which are associated with conversations and informal discourse rather than academic writing. Examples of vague nouns include words like *human, people, stuff* which are used to refer to objects and/or events without giving a specific definition. Hinkel attributes learners' use of such nouns to their lack of knowledge of more advanced vocabulary (Hinkel, 2002:83). Out of the six verb semantic classes Hinkel investigated in her study, namely public, private, suasive, logical-semantic,

verbs of expecting/wanting/ tentative, and seem/appear verbs, NNS use predominantly three classes: public, private, and expecting/tentative verbs. Public verbs refer to actions that can be observed publicly, such as *acknowledge, agree, protest, offer, speak;* private verbs refer to intellectual status, such as *accept, discover, feel, suppose;* expecting/tentative verbs refer to future time, such as *desire, plan, want.* Within each class of verbs, it was noted that learners use lexically simple and commonly used verbs, such as *argue, believe, decide, say,* more frequently than other complex less common verbs, such as *acknowledge, advocate,* and *certify* (Hinkel, 2002:104-107).

2.7 Research aims and research questions

In Study 1, of this thesis, the researcher attempts to investigate the written production of Arabic speaking learners of English which is, to the best of the researcher's knowledge, an under-researched area. A corpus-based approach to collocations is used to facilitate the analyses of a large amount of data and the identification of quantitative and qualitative differences in the use of two types of lexical collocations, verb-noun and adjective-noun collocations. Study 1 aims to ascertain which of these two types of collocations would be more fruitful for identifying aspects of academic writing by Arabic learners of English that would benefit from improvement.

The results of Study 1 reveal that verb-noun collocations are more difficult for Arab learners than adjective-noun collocations. Therefore, driven by the interest in collocations, lexical and grammatical, the understanding of the role of the verb as the most important element in a given structure based on which many lexical choices are made, and influenced by the general shift in literature towards lexico-grammatical patterns, Studies 2 and 3 of this thesis adopt a novel approach which analyses the syntactic and the semantic features of verb complementation clause structures and the verb-noun collocations embedded within these clause structures.

Studies 2 and 3, attempt to make several contributions to research on verb-noun collocations and verb complementation clause structures. It adopts an innovative usage-based approach that aims for a comprehensive contrastive interlanguage analysis (CIA) of the use of verbs in selected clause structures in academic writing produced by expert and novice writers in both English and Arabic. This study is also innovative in the field of CIA in terms of the target languages. Learner corpus is not only compared to that of native

speakers but also to corpora of experts writing which represent 'reference language varieties' (RLV) advocated by Granger (2015:17). Furthermore, this study is one of the few studies that extensively explore advanced Saudi learners' use of verb complementation clause structures and VN collocations through the analysis of a corpus of learners' authentic academic writing.

The main purpose of this study is to explore advanced Saudi learners use of selected verb complementation clause structures and VN collocations that occur within the frame of these clause structures and the difficulties that learners may encounter. The register under investigation is that of academic writing in the discipline of applied linguistics. The study also aims to investigate the influence, positive or negative, of learners' first language, Arabic, on the use of English clause structures. It is hoped that the outcomes of this study would increase the understanding of the register of academic writing, specifically the use of verb complementation clause structures. Because academic writing produced by advanced Saudi learners of English is studied here, this study also hopes to provide some pedagogical suggestions and implications that could help this target group write better in academic English. In more specific terms, the objectives of Studies 2 and 3 of this thesis are as follows:

1) To investigate the similarities and the differences between Arabic and English in the use of the target clause structures and VN collocations in the register of expert academic writing in applied linguistics;
2) To investigate the similarities and the differences in the use of the target clause structures and VN collocations in the register of novice academic writing in applied linguistics produced by novice native and novice non-native writers, the latter being advanced Saudi learners of English;
3) To explore the similarities and the differences in the use of the target clause structures and VN collocations by expert writers vs. novice writers.
4) To identify areas of difficulties for advanced Saudi learners of English and trace the influence of the first language, positive or negative, on learners' use of English clause structures and VN collocations.
5) To provide pedagogical implications for the teaching and learning of verb clause structures and VN collocations in English based on the findings of the study.

In order to achieve these objectives, two sets of questions are put forward. The first set is related to the use of verb complementation clause structures and verb-noun collocations in Arabic and English expert academic writing in the field of applied linguistics. This set includes:

1) Which of the selected clause structures are frequently used in expert academic writing in both languages? What are the similarities and the differences in the use of these clause structures, if any, between the two languages?

2) What are the most frequent verbs used in each of the selected clause structures in both languages? What are the syntactic and the semantic representations of the verbs used in each of the selected clause structures in both languages?

The second set of questions is related to the analysis of the use of verb complementation clause structures and verb-noun collocations in the academic writing produced by two groups of novice writers - native British students of English Language and Linguistics and counterpart novice non-native Saudi advanced learners of English. These questions are as follows:

1) Which of the selected clause structures are frequently used in academic writing by novice writers? Which of these selected clause structures do advanced Saudi learners overuse or underuse as compared to novice native speakers?

2) What are the most frequent verbs used in each of the selected clause structures by both groups of novice writers? What are the syntactic and the semantic representations of the verbs used in each of the selected clause structures by both groups of novice writers?

3) Is there a relationship between advanced Saudi learners' use of clause structures and their first language, Arabic?

Chapter 3: Study 1

This chapter briefly presents a description of the first study conducted in the journey of this thesis which represents the phraseological approach to the study of lexical collocations. The chapter starts with the aims and research questions of Study 1. Then, it provides a description of the methodology for the compilation of native and non-native corpora and the procedures followed for the analysis of these corpora. The results of this study are discussed in Section 3.3. At the end of the chapter, the limitations and the implications of Study 1 are presented.

3.1 Study 1 aims and research questions

Study 1 aimed to explore the use of two types of lexical collocations, verb-noun and adjective-noun by advanced Arab learners of English in order to find out which of these two types is more difficult for Arab learners of English. To achieve this aim, a learner corpus was compiled and then compared to a native speakers' corpus. The researcher explored patterns of learners' use specifically whether learners tend to overuse, underuse or misuse these two types of collocations in comparison to novice native speakers. The researcher also investigated the misused collocations and attempted to discover the causes behind this misuse and trace the influence of the learners' first language, Arabic. The study sought to answer the following research questions:

1) What are the differences in the use of adjective-noun and verb-noun lexical collocations between Arab learners of English and native speakers of English? Do Arab learners tend to overuse, underuse or misuse certain collocations? And if so, which ones?

2) Is there a relationship between learners' misuse of lexical collocations and their first language (Arabic)?

3.2 Methodology

3.2.1 Learner Corpus- TEEP-ArSL

The learner corpus for study 1, the TEEP of Arabic speaking learners (TEEP-ArSL), was compiled from argumentative essays written by Arab learners in the writing section of the Test of English for Educational Purposes (TEEP). TEEP is a university entrance test of

proficiency for students whose first language is not English. This test was chosen as the source of data for three reasons. It is a standardised test, and students' writing has been marked and moderated by two raters, which maximises the reliability of the evaluation of students' proficiency level. Furthermore, the writing is mainly argumentative and since the ultimate goal of this study is to suggest strategies for improving learners' lexical proficiency in English at advanced level (B2 and above), it is reasonable to focus on a register which is required from learners in order to enter British Higher Education.

The sample included written texts produced from 2013 to 2015 by 130 speakers of Arabic as their first language. The essays vary in length from 300 to 500 words and cover topics such as: teleworking, oil as a source of energy, tourism, long working hours, local and international crimes, and happiness. The researcher, with the help of the head of the ISLI (International Study and Language Institute at the University of Reading) testing team, carefully checked the participants' nationalities to ensure they are from Arabic-speaking countries. The sample includes Arabic speakers from Saudi Arabia, Kuwait, Qatar, Oman, Bahrain, Libya, Iraq and Jordan. To control for the examinees' proficiency level, only essays that scored 5.5 and above were included. These essays are considered higher intermediate to advanced, that is equivalent to B2 – B1 level in the common European Framework of Reference for languages (CEFR) (see Table 3-1 for the TEEP scores and their CEFR equivalent).

Table 3-1: TEEP scores and CEFR level equivalent

TEEP score	CEFR level
8 – 9	C2: Mastery
7 -7.5	C1: Effective Operational Proficiency
6 – 6.5	B2: Vantage
5 -5.5	B1: Threshold
4 - 4.5	A2: Waystage
3.5	A1: Breakthrough

After collecting 130 essays that met the above criteria, the researcher manually transcribed the essays. Care was taken not to correct any of the learners' errors, so as to create a raw learner corpus, i.e. learners' original plain texts with no extra features added (Granger, 2002). Quotes in the learners' essays were deleted and replaced by '0', so that

quoted data would not affect the analysis of learners' production. The transcribed essays were uploaded to Sketch Engine and compiled into a corpus, TEEP- ArSL, which consists of 56,688 tokens and 4,610 types.

3.2.2 Native Speaker Corpus- LOCNESS-A-Level

The NS control corpus that was used in Study 1 is a subsection of the Louvain Corpus of Native English Essays (LOCNESS). This corpus is compiled by the Centre for English Corpus Linguistics (CECL) at the Université catholique de Louvain in Belgium. It contains four components: argumentative essays written by British A-Level students, essays written by British university students, argumentative essays by American university students and literary-mixed essays by American students. The texts of the corpus include examination papers, timed essays and free essays. The length of essays is around 500 words. The texts cover a wide range of topics such as water pollution, nuclear power, sex equality, violence and animal testing. The total number of words in the corpus is 324,304 and the number of types is 19, 941.

The first component of the LOCNESS which is the British A-level, henceforth referred to as LOCNESS-A-Level, was chosen as a comparable corpus in this study for various reasons. Its language-related features are similar to the learner corpus. LOCNESS-A-Level includes 500-word written argumentative essays, which is comparable to the length and genre of the essays in TEEP-ArSL, 300 to 500-word argumentative essays. Essays in LOCNESS-A-Level are taken from mock-exam papers and the essays in TEEP-ArSL are exam papers. Furthermore, writers of both corpora are not expert writers and the topics are not academic or specialised but are general argumentative essays prompting the writers to express their opinions. Topics of the LOCNESS-A-Level include: transport, parliamentary system, fox hunting, boxing and computers. The size of this component is also relatively comparable to the size of the learner corpus as it consists of 66,545 tokens and 6.838 types while TEEP-ArSL consists of 56,688 tokens and 4,610 types.

There are some differences between the two corpora in the topics discussed and the writers' background. However, finding an available native speaker corpus that is exactly comparable to the learner corpus is a very difficult task. Given the fact that the source of data used in this study is a proficiency test, it is very difficult to have native speakers sit for an actual proficiency level test. One could ask a group of native speakers to produce the same kind of essays but given the time constraints of this phase of the research, this was

not feasible. This problem of finding a comparable native speaker corpus has been faced by researchers who conducted similar studies. For example, Granger (1998a) has acknowledged this problem of finding an exact comparable corpus and noted that most control corpora can be criticised as not being a direct equivalent. Having said that, Granger (1998a) believes that it is important for researchers to be aware of these limitations and make choices as best as possible based on the type of investigation of their study and the data which are available. Hence, the researcher decided to use a data set which is not fully, but in many ways comparable, in terms of the genre and the text size, to the learner data.

3.2.3 Procedures of study 1

For Study 1, the procedures used by Laufer and Waldman (2011) were adopted with some modifications. Instead of the Wordsmith Tools used by Laufer and Waldman (2011), Sketch Engine (Kilgarriff et al., 2014) is used. Sketch Engine was preferred in this study as the basic search tool because it allows the upload of the researcher's own corpora and has many advanced functions, such as the Corpus Query Language (CQL), which enables users to input complex search queries for specific structures or collocations. Furthermore, Sketch Engine has the advantage of automatically POS tagging the corpus as it is uploaded. The main steps for the data analysis were as follows:

1. Using the word list function in Sketch Engine (Kilgarriff et al., 2014), the most frequent nouns were retrieved from the two corpora: the LOCNESS-A-Level (NS corpus) and the TEEP-ArSL (learner corpus). The cut-off point was 20-occurrence, that is only nouns that occurred 20 times and more in the two corpora were included. One-hundred-and-three nouns were extracted from the LOCNESS-A-Level that occurred more than 20 times, and 90 nouns from the TEEP-ArSL. Thirty-four nouns, found in common in the two corpora were used as the base of the investigation, see Table 3-2 for these nouns, their raw and normalised frequencies, per 100,000.

Table 3-2: Common nouns in the two corpora and their frequencies

	Common Nouns	LOCNESS-A-Level		TEEP-ArSL	
		Raw Freq	Norm Freq	Raw Freq	Norm Freq
1	time	399	703.85	64	96.18
2	work	275	485.11	56	84.15

	Common Nouns	LOCNESS-A-Level		TEEP-ArSL	
		Raw Freq	Norm Freq	Raw Freq	Norm Freq
3	life	203	358.10	67	100.68
4	problem	143	252.26	74	111.20
5	example	143	252.26	68	102.19
6	society	134	236.38	35	52.60
7	world	120	211.69	75	112.71
8	number	117	206.39	69	103.69
9	family	104	183.46	24	36.07
10	country	101	178.17	55	82.65
11	government	95	167.58	78	117.21
12	money	85	149.94	135	202.87
13	effect	69	121.72	37	55.60
14	use	57	100.55	33	49.59
15	issue	57	100.55	27	40.57
16	future	57	100.55	25	37.57
17	way	53	93.49	73	109.70
18	day	50	88.20	29	43.58
19	production	42	74.09	25	37.57
20	responsibility	41	72.33	26	39.07
21	need	38	67.03	20	30.05
22	increase	36	63.51	26	39.07
23	part	33	58.21	32	48.09
24	situation	32	56.45	20	30.05
25	place	31	54.69	26	39.07
26	food	30	52.92	33	49.59
27	reason	28	49.39	35	52.60
28	technology	26	45.87	39	58.61
29	demand	25	44.10	30	45.08
30	case	20	35.28	60	90.16
31	fact	23	40.57	42	63.12
32	sport	22	38.81	103	154.78
33	power	20	35.28	25	37.57
34	opinion	20	35.28	24	36.07

2. Verb-noun collocations of these 34 most frequent nouns were retrieved from the two corpora using the CQL function available in Sketch Engine. Two strings were used to retrieve these collocations. For example, for the noun *time*, the following instructions were used:

    [tag="V."][word="time"]
    [tag="V."][word=".*"][word="time"]

The second string was used to detect any verb-noun collocations that might contain any modifier before the noun. Out of the list created as a result of these two strings, the copular *be* was excluded. This decision was justified by the fact that the combination of the verb *be*+ noun does not create a collocation but rather a free combination.

The same procedure was followed to retrieve the adjective-noun collocations of the same 34 most frequent nouns, except that one string was used to retrieve the collocations. For example, for the noun *time*, the following instruction was used: [tag="JJ.*"][word="time"]

For convenience of automatic retrieval, the focus was mainly on attributive adjectives, such as *long time*. Predicative adjectives, such as the time is short, although they can form a collocational relation with nouns were not considered. Determiners and quantifiers, such as *most, more, many* and *such*, were excluded from the data as they along with the following nouns form free combinations rather than collocations.

The extracted combinations, both verb-noun and adjective-noun, were lemmatised and grouped into types. For example, the combinations *have responsibility*, *has responsibility*, *had responsibility*, *have responsibilities* and so on are grouped under one type, *have responsibility*. This step facilitated the following steps of the analysis and the comparison between the types and the tokens.

3. The extracted combinations were checked in two dictionaries: *The BBI Dictionary of English Word Combinations* (Benson, Benson & Ilson, 1997) and *the Oxford Collocations Dictionary* (McIntosh, Francis & Poole, 2009). These two dictionaries were chosen because they represented the two approaches to collocations and they were used in previous studies. The *BBI* represented the phraseological approach as it is a non-corpus-based dictionary and it was consulted in studies by Nesselhauf (2003), Laufer and Waldman (2011) and Marco (2011). The *Oxford Collocations Dictionary* represented the frequency-based approach as it is based on Oxford English Corpus (http://www.oxforddictionaries.com/words/the-oxford-english-corpus) and it was consulted by Nesselhauf (2003) and Marco (2011). If the combination was listed as a collocation in one of the dictionaries, it was noted as such. If the combination was not listed in any of the dictionaries, it was searched

for in the British National Corpus (BNC-BYU) (Davies, 2004). The frequency of occurrence of the collocation and the Mutual Information (MI-score) were taken into consideration. MI-score is a frequency measure of the strength of a collocation that operates on a scale that does not have a theoretical minimum and maximum. The larger the value the more exclusively the two words are associated (Gablasova et al., 2017: 163). If the score is 3 or higher, it reflects a semantic bonding and it can be taken to be significant (Hunston, 2002: 70-71). Thus, if the combination had a score of 3 or higher, it was noted as a collocation. If it had a lower score than 3, the collocation was disregarded, and it was considered a free combination. Combinations that had a very low frequency of occurrence (1 or 2 occurrences) and combinations that were not found at all in the BNC were listed, along with an example taken from the concordance lines, to be checked by two native speakers.

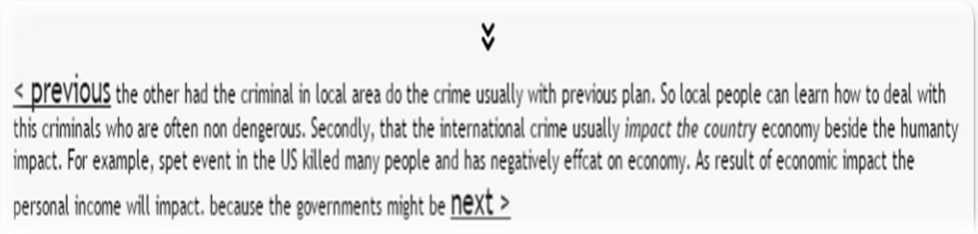
The two raters' agreement is important to ensure that their judgment is not affected by any factor related to one of the raters, such as their background or personal experience (Gwet, 2014: 4). Agreement on the acceptability or unacceptability of the combination maximises the inter-rater reliability (IRR).

Two lists were created: one for verb-noun collocations and it consisted of 28 combinations from the LOCNESS-A-Level and 39 from the TEEP-ArSL, while the other list was for adjective-noun collocations and it consisted of 42 combinations from the LOCNESS-A-Level and 36 from the TEEP-ArSL. The raters, who were teachers of English as a Second Language, were asked to read the lists and judge the given collocations as acceptable or unacceptable. IRR was calculated using the Cohen's Kappa statistical measure. This measure was used in this study because it is a preferred measure to assess IRR for fully-crossed nominal data, such as the data at hand (Hallgren, 2012). The IRR for the raters of verb- noun collocations was found to be Kappa = 0.572, which is considered a moderate agreement (Landis & Koch, 1977). For the adjective-noun collocations, the IRR was found to be Kappa= 0.612, which is considered a substantial agreement. As clear from the Kappa, the IRR did not reflect perfect agreement, and that was why, for the items on which the two raters disagreed, a third rater, who was also a teacher of English as a Second Language, was asked for final judgement. Combinations which the raters agreed on their acceptability were disregarded while those that the

raters considered unacceptable were considered misused collocations and were listed as such. As a result of this step, two excel sheets were created which included the correct collocations and the misused combinations of both types, verb-noun and adjective-noun.

The term 'misused collocations' rather than 'deviant', 'errors' or 'incorrect collocations' was used in this study to refer to the combinations that were marked as unacceptable by the raters. That is because the combinations that were classified under this group were not necessarily incorrect collocations, but, in the given examples, the use of that combination was inappropriate. For example, the combination *impact the country* was considered inacceptable because the subject in the example sentence (see Figure 3-1) was *crime*, and *crime*, according to the raters, could not *impact the country* but usually *had an impact*.

Figure 3-1: context of [*impact*] [*country*]



4. A further step was taken to analyse misused collocations by the learners and check whether they were due to the influence of their L1. To maximise inter-rater reliability, two native speakers of Arabic, who also have a good command of English, were asked to judge if these collocations were a result of literal translation of an Arabic equivalent. The two raters were the researcher and a Saudi PhD student of applied linguistics. Two lists were created: a list of 26 verb-noun combinations and another list of 27 adjective-noun combinations. The IRR of the two raters' evaluations was also calculated using the Cohen's Kappa measurement and it was found to be Kappa= 0.839 for the verb-noun combinations list and Kappa=0.834 for the adjective-noun combinations list which may be considered, according to Landis & Koch (1977), an almost perfect agreement. It was planned to compare the misused combinations to their Arabic homologues in the Arabic corpus: King Saud University Corpus of Contemporary Arabic (KSUCCA) (Alrabiah, Al-Salman, &

Atwell, 2013). However, this step was not feasible due to the fact that some English nouns, for example *opinion,* would have more than one equivalent in Arabic, such as */ra?i/, /ara?/, /wʤhat naðˤar,/fikrah/, /alfikrah/* and so on. Furthermore, the variation in the Arabic sentence structure made it difficult to detect the target combinations using the CQL function. Thus, the two Arab native speakers' judgement which had a relatively high inter-rater reliability was considered sufficient. KSUCCA and Google Translate (https://translate.google.co.uk/) were used as helpful tools for the qualitative analysis.

5. For comparison purposes, normalised frequencies were calculated. The base of the normalisation in study 1 was 100,000. This base seemed to be appropriate because it is larger than the sizes of the two corpora (LOCNESS-A-Level =66,545 and TEEP-ArSL=56,688).

   Type/token ratio (TTR) calculated in its simple form varies in accordance with the length of the text, that is, longer text results in lower TTR because function words tend to be repeated in longer texts (Baker, 2006). In this research the same effect may occur when comparing lists of collocations with different lengths. Therefore, to account for lexical diversity, the Guiraud's index, which corrects to some extent for the number of tokens, is used in this study (Daller, 2010; Guiraud, 1954; Tidball & Treffers-Daller, 2007). The Guiraud's index is calculated by dividing the total number of types by the square root of the total number of tokens: Types/ √ tokens.

   To check for the statistical significance of the frequencies, Log-likelihood test (LL) is used. The LL is a preferred test, rather than the Chi-square, for corpus studies. The LL value is considered significant at the level of $p<0.05$ when it is greater than 3.84 and the degree of freedom for all calculations used in this study is $df=1$, as we are comparing two corpora (http://ucrel.lancs.ac.uk/llwizard.html). For question 2, the misused collocations have been classified into two groups: misused collocations, and misused collocations caused by L1 transfer. Percentages of the misused collocations in the two groups are calculated.

3.3 Results and discussion

The results of the analysis of the use of verb-noun collocations is represented first, followed by the results of the analysis of the use of adjective-noun collocations.

3.3.1 Verb-noun collocations

The results of the data analysis of the use of verb-noun collocations revealed that Arab learners used significantly more tokens of correct verb-noun collocations than native speakers (393 normalised tokens in TEEP-ArSL compared to 201 normalised tokens in LOCNESS-A-Level). However, when normalised types were compared the difference between the two corpora was not significant (115 in TEEP-ArSL compared to 98 in LOCNESS-A-Level). These results seem to contradict the findings of some elicitation studies on Arab learners, such as Farghal and Obaidat (1995) and Hussein (1998), which concluded that learners tend to avoid unfamiliar collocations, and the findings of Laufer and Waldman (2011), who concluded that learners at different levels of proficiency produce far fewer collocations than native speakers. It was evident in this study that Arab learners at an advanced level of proficiency did not avoid the use of verb-noun collocations but produced more verb-noun collocation tokens and a comparable number of verb-noun collocation types to native speakers.

The Guiraud's index of the type/ token ratio revealed that although advanced Arab learners used a comparable number of correct verb-noun collocation types to native speakers, the lexical diversity of these correct collocations was less as the type/token ratio was lower (5.8 to 6.91), which indicates that learners tend to frequently repeat the collocations they use. A number of collocation types in TEEP-ArSL had a high frequency of more than 30, for example, *spend time* (38 occurrences), *solve the problem* (30 occurrences), *have time* (23 occurrences) and *earn money* (14 occurrences) (see Appendix 2). These findings were in line with the results of Altenberg and Granger (2001), Nesselhauf (2005) and Jaworska et al. (2015) which showed that L2 learners overuse certain repetitive expressions. Learners' overuse of these more frequent collocations can be explained by the so-called "teddy bears" effect (Hasselgren, 1994; Nesselhauf, 2005). Learners tend to rely on a smaller number of verb-noun collocations (lexical/collocational teddy bears) which are commonly used in everyday life because these teddy bears represent safe choices.

Advanced Arab learners misused 28.6 % of the verb-noun collocations identified. The influence of the learners L1 could be traced in 42.31% of the misused verb-noun collocations. This high percentage appears to confirm the results of Nesselhauf (2003), who reported a higher percentage of 56%, and Laufer & Waldman (2011), who reported that the percentage of deviation influenced by Hebrew ranges from 44% to 64.5% across three levels of proficiency. The salient influence of learners' L1, Arabic, on their production of verb-noun collocations in Study 1 also seems to confirm the results of studies by Farghal and Obeidat (1995) and Hussein (1998) which showed that most Arab learners' incorrect responses to test items were due to the use of inappropriate L1 translation. For example, in Study 1, the misused collocations *lose their future** and *solve the issue** were the result of a literal translation from Arabic: */yaχsaru mustaqbalah/(lose his future)* and */yaħulu ʔalmsʔlah/(solve the question).*

To support the evidence for L1 transfer, examples of misused verb-noun collocations that were judged by the Arabic raters to be a result of L1 transfer (see Table 3-3) were checked and compared to examples listed in other studies on the use of verb-noun collocations by learners with a different L1. None of the eleven examples listed below were produced by German learners in Nesselhauf (2005), nor by Chinese learners in Yanjuan (2014).

Table 3-3 : misused verb-noun collocations in the TEEP-ArSL as a result of L1 transfer

	Misused verb-noun collocations		Misused verb-noun collocations
1	coordinate time	7	secure the need
2	damage the country	8	reduce the situation
3	reach the society	9	increase the situation
4	decrease negative effect	10	stop the situation
5	lose their future	11	worries the world
6	peak their production		

3.3.2 Adjective-noun collocations

The results of the analysis of advanced learners' correct use of adjective-noun collocations showed that the frequency of the normalised tokens of adjective-noun collocation was significantly higher in the learner corpus when compared to native speakers (609 normalised tokens in the TEEP-ArSL to 403 normalised tokens in the LOCNESS-A-

Level). However, the comparison between the frequency of the correct adjective-noun collocation normalised types showed no statistically significant difference in the use of this type of collocations, between advanced Arab learners (229) and native speakers (231). This result seems to confirm the results of Siyanova and Schmitt (2008) who concluded that there is no significant difference in the use of adjective-noun collocations between native speakers and advanced learners of English. 24.7/ 20

The Guiraud' index for the type/token ratio of adjective-noun collocations showed that the lexical diversity of the correct adjective-noun collocations used by Arab learners was lower than that of native speakers (9.27 to 11.55). Arab learners tend to repeat certain commonly used adjective-noun collocations such as *long time* (40 occurrences), *social life* (20 occurrences), *positive effect* (12 occurrences), *best way* (10 occurrences) and *daily life* (9 occurrences). These results can also be explained by the so-called "teddy bears" effect reported in previous research (Hasselgren, 1994; Nesselhauf, 2005).

Arab learners misused 17% of adjective-noun collocations. Compared to the percentage of misused verb-noun collocations (28.6 %), this percentage is relatively smaller. Although Arab learners misused adjective-noun collocations less frequently than verb-noun collocations, 70.37% of the misused adjective-noun collocations were due to the influence of Arabic. It seems that learners tend to rely heavily on their first language when producing adjective-noun collocations and as a result of the use of an inappropriate synonym or incorrect L1 translation, they produce collocations that may be considered unacceptable by native speakers of English. For example, *respectful places** is a result of literal translation of the combination */amakin muħtaramah/(respectful places)* and *passive effect** appears to be a result of the use of an inappropriate near synonym. This appears to confirm the results of Farghal and Obeidat (1995) and Hussein (1998) who considered the negative transfer from learners L1, Arabic, to be one of the most important factors that lead to learners' misuse of collocations.

To conclude, the results of Study 1 demonstrated that Arab learners misuse 17% of adjective-noun and 28.6% of verb-noun collocations. This larger proportion of misused verb-noun collocations compared to adjective-noun collocations indicated that Arab learners encounter greater difficulty with verb-noun collocations. These results were consistent with the results of Siyanova and Schmitt (2008) which indicated that adjective-noun collocations are easier for learners than verb-noun collocations. These results also

confirmed the findings of Javid and Umer (2014: 176) and Al-Khairy (2013: 5-6) which report Arab learners' difficulty with appropriate use of verbs.

The identification of verb-noun collocations as a more difficult type of collocations for Arab learners as compared to adjective-noun collocations encouraged the researcher to focus on this type of collocations and conduct a more detailed study. Therefore, the rest of the thesis focuses on verb-noun collocations rather than adjective-noun collocations.

3.3.3 Limitations of the study

Although Study 1 revealed some interesting results in relation to advanced Arab learners' use of lexical collocations, it had a number of limitations.

The focus on VN lexical collocations led to the exclusion of many grammatical collocations, especially verb+ particle. Due to the time constraints of Study 1, this type of collocations was not discussed. Furthermore, most of the VN collocations retrieved in Study 1 are examples of the transitive clause structure. Learners' use of single-word and multi-word verbs in other clause structures is an interesting area of research that could reveal valuable information.

Another challenge that faced the researcher in Study 1 was the definition of collocations. This challenge was also identified by Nesselhauf (2003). The distinction between collocations and free combinations was not easy to define. In Study 1, collocation dictionaries, the BNC, and MI- score were considered in the process of defining collocations. Native speakers' judgment was also sought. The use of statistical tests, such as MI and T-score, available through different corpus tools, such as Sketch Engine and BNC-BYU, were potentially helpful; but the researcher had to use them with caution because each test yields different results. While the MI favours less frequent content words, the T-score favours high frequency function words. So, using these tests as the only criterion for judging lexical collocations might not be reliable. Seeking native speakers' judgement, as raters of the acceptability or unacceptability of the target collocations was deemed to support the use of statistical measures. However, the judgement was subject to the raters' age, their background, their language exposure and professional experience because what is acceptable in one field may not be viewed as acceptable in another.

Study 1 focused on investigating the influence of L1 negative transfer on learners' misuse of collocations. However, exploring the influence of L1 positive transfer on learners' use of correct collocations can be helpful. It could be useful for teachers and

syllabus designers in the Arab world to know about the collocations that are common in Arabic and English and which may lead to correct use of the English collocations.

3.4 Methodological Implications of Study 1

Based on the results and the limitations of Study 1, the researcher made a number of decisions to improve the methodology of the proposed research, i.e. Studies 2 and 3. These included:

1) Following the general shift to the study of collocations from the focus on the one-dimensional association of words, to a more holistic view that emphasises the interdependency between lexicon and grammar, a usage-based approach that emphasises the association between meaning and grammar is adopted. Rather than the focus on the co-occurrence of lexical items, this approach focuses on the co-occurrence of the lexical item, specifically the verb and the following noun phrase, within the frame of verb complementation clause structures. The decision to adopt this approach is justified by many reasons. First, the choice of verb in these clause structures has an element of unpredictability which may pose difficulty for learners similar to that of collocations. In addition, such approach provides a more comprehensive description of language based on the interdependency between grammar and lexicon. Furthermore, this approach facilitates the study of both single and multi-word verbs that may fill in the verb slot in the clause structures.
2) A review of two approaches to the study of lexico-grammatical patterns, namely Pattern Grammar and Construction Grammar (Sections 2.2. and 2.4), revealed some limitations of these two approaches. Whereas the description of patterns in pattern grammar is too detailed to be used for a comprehensive description of the use of verb and its complementations, the description provided by Construction Grammar is too limited. Therefore, Quirkian clause structures (1985) are used as the base of the syntactic identification of the lexico-grammatical patterns of verb use in academic writing. Semantic Frames and FrameNet are referred to for the semantic analysis in both Study 2 and Study 3.
3) The contrastive interlanguage analysis in Study 1, compares a corpus of non-native Arab learners to a reference corpus of native speakers' writing. However,

in Studies 2 and 3, a variety of reference corpora are included, two corpora of expert writing in English and in Arabic as well as a corpus of novice native speakers writing. This model of CIA provides different reference points against which learner corpus is investigated (Granger, 2015: 17). It also facilitates the identification of the similarities and the differences between the academic writing of two groups of expert writers and between expert and novice writers in relation to the use of verb and its complementations.

4) While Study 1 focuses on learners' misused collocations to trace the influence of their first language (negative transfer), Studies 2 and 3 attempt to identify the similarities and the differences between the learners' first language, Arabic, and the target language, English, in relation to the use of verb complementation clause structure and the choice of verb therein. This step is particularly useful in the area of Second language acquisition and first language transfer because of the significant role of similarities between the L1 and the L2 in reducing the time and effort of the learning process (Ringbom, 2007: 1). Ringbom (2007) also believes that in their search to facilitate the learning of a target language, learners focus on language similarities not language difference. Adopting a CIA model, in Studies 2 and 3, makes it possible to look at features that are shared between the L1 and the L2 and other features that are specific for one language and discover patterns of positive and negative transfer in relation to verb complementation structures and VN collocations.

3.5 The importance of the proposed approach

In this section, the researcher would like to reflect on the importance and usefulness of the proposed approach which draws on Quirkian clause structures for the syntactic analysis and Frame Semantics for the semantic analysis of verb-noun collocations within the frame of verb complementation clause structures as compared to the phraseological approaches. An exemplar verb, *make,* is selected to demonstrate the usefulness of the approach. This verb is chosen because it is used in most of the selected clause structures for Studies 2 and 3.

In Study 1, the researcher used two search strings to retrieve verb-noun collocations. Using the same two strings, [lemma="make"][tag="N.*"] and

[lemma="make"][tag=".*"][tag="N.*"], to retrieve the VN collocations of the verb *make* results in the following tables:

	[lemma="make"][tag="N.*"]	
1	make sense	18
2	make use	16
3	making use	9
4	makes use	7
5	making sense	5
6	making errors	5
7	makes sense	5
8	make judgments	5
9	make errors	5
10	made use	5
11	making mistakes	4
12	make promises	4
13	make claims	4
14	made sense	4
15	made pancakes	4

	[lemma="make"][tag=".*"][tag="N.*"]	
1	made an effort	4
2	making language errors	3
3	make the case	3
4	make any claims	3
5	make a distinction	3
6	made a mistake	3
7	made a game	3
8	making a statement	2
9	making a game	2
10	making a claim	2
11	makes this constraint	2
12	makes the statement	2
13	makes the construction	2
14	makes little sense	2
15	makes an appeal	2

This methodology provides some useful information for learners. Lexical collocations, such as *make mistakes*, *make an effort,* and *make promises* are identified. However, the methodology employed does not provide sufficient understanding of the use of the verb *make* in larger structures nor does it say much about the possible particles that may combine with this verb to form multi-word verb combinations.

The proposed methodology is based on the idea of exploring all occurrences of the verb *make* and classifying its use according to the syntactic structures in which *make* may occur. Then, a semantic analysis based on Frame Semantics is conducted to understand the semantic roles involved in the identified clause structures for the verb *make*, especially when its used as part of a multi-word verb. An example of the results based on the proposed methodology would include information about the use of *make* in the phrasal type 2 clause structure, e.g. *make up*, in the prepositional type 3 clause structure, e.g. *make use of, make sense of, make reference to,* and in the phrasal prepositional type 2 clause structure *made up of* (16). The semantic analysis would provide information about the meaning of the combination and the semantic frame it activates. For example, *made up of* activates the semantic frame of 'creating' which involves a creator, a created entity and components

used for the creation mentioned after the preposition *of*, for example, *the acronym is made up of the initial letters of the dimensions.*

Furthermore, the approach adopted for Studies 2 and 3 is concerned with the association between the lexical items and the clause structure at the paradigmatic level. It provides information regarding the possible variations that may occur in a given clause structure. For example, in the case of the verb *make* and its occurrences in prepositional type 3 clause structure, e.g. *make use of,* information about the use of *make* in other prepositional type 3 combinations, e.g. *make sense of, make fun of* is also provided. Additionally, the use of other verbs in the same clause structure, such as *take care of, give attention to, play a role in* is presented. The measure of 'faithfulness' (Römer et al., 2015) is used to decide on this association between the verb and the clause structure in which it is used, see Section 4.1.4 for detailed explanation of this measure. This measure normalises the occurrences of a verb in a given clause structure by the total occurrences of that verb per hundred which reflects the proportion of the verb's occurrences in that clause structure. For example, if a verb has a total frequency of 1000 and it is used in the transitive clause structure in 900 occurrences, this means that this verb is highly associated and 'faithful' to the transitive clause structure as it occurs in this clause structure in 90% of its total occurrences. The measure of faithfulness of the verb to the clause structure is useful because it provides information about verbal variation in a clause structure and about the most preferred type of complementation for a given verb.

Compared to the previous methodology used in Study 1 which results in a report of the one-dimensional use of the verb along with a following noun, the methodology proposed for Studies 2 and 3 would reveal much more information about the association of the verb with the clause structure and highlight fine-grained differences in the use of certain verb-noun collocations.

Chapter 4: Research Methodology for Studies 2 and 3

The methodological implications of Study 1 were implemented in the methodology for Studies 2 and 3 in order to achieve an increased understanding of the choice of the verb and the following complementation within the frame of clause structures. This chapter gives a detailed description of this methodology.

4.1 Methodology for Study 2

The purpose of this study is to investigate the use of the verb and its noun phrase complements by systematically exploring the syntactic and the semantic frames of the most frequent verbs in academic writing in English and Arabic. The analysis of expert writers' data in both languages serves two functions. It helps to understand similarities and differences between the two languages, and hence could pinpoint to areas of positive and negative transfer. It is also intended to be used as a benchmark for further comparison of novice writers' use of verbs and verb complementation clause structures in academic writing to better understand which ones a learner of English with Arabic L1 might find difficult or easy. To achieve these goals, appropriate data sources had to be identified in the first instance.

The researcher aims to help advanced Saudi learners improve their English academic writing and in order to have a good sample of general academic writing, sample texts from a variety of genres and disciplines need to be gathered. However, this task is impossible within the limits of a PhD thesis. So, in order to narrow the focus of this study to a homogenous disciplinary sample, texts from applied linguistics were selected. The field of applied linguistics represents a good source of data because, as mentioned in Section 1.1, it is one of the fields in Saudi Arabia where learners need to write academically in English at advanced levels. Since the researcher aims to help advanced learners improve their academic writing, it is practical and useful to focus on a discipline where this type of writing is most required. Consequently, decisions on the data for the reference corpora of expert writing were made to match the learner data which focuses on the discipline of applied linguistics.

It is commonly recognised that published research articles are the models of academic writing (e.g. Swales, 1990) and thus the decision was made to investigate academic articles published in recognised journals in the English and Arabic academic

communities. As explained in Section 1.1, the choice of published journal articles is justified by the fact that the purpose of this study is not to impose the norms of published work on learners, but rather to describe how verbs are used in different clause structures in the writing of the experts which provides a good model of lexico-grammatical choices and possibilities.

Thus, for Study 2, the researcher compiled two corpora of academic published writing in English and Arabic in the field of applied linguistics. The researcher tried to make the two corpora as representative and comparable as possible. The following sections discuss the choice of journals and articles.

4.1.1 Academic English Corpus (AEC)

The researcher used the University of Reading Library subscription to access a variety of journals in the field of English Language and Applied Linguistics. The researcher attempted to make the corpus as representative of the field of English language and applied linguistics as possible by including journals that focused on a variety of applied linguistics subject areas including language teaching, sociolinguistics, discourse studies as well as theoretical linguistics. These areas are widely covered in teaching English Language and Applied Linguistics at the university level. Given the availability, nine English journals were selected: Applied Linguistics, Journal of Linguistics, TESOL, Discourse and Society, Journal of Sociolinguistics, Language and Society, English Language and Linguistics, Discourse and Communication, Studies in Language. From each journal 10 to 15 articles were chosen, care was taken to include articles with both quantitative and qualitative methodologies and to select the most recent ones. Table 4-1 presents the journals, the number of articles from each journal, their dates, and word size.

The selected articles were transferred from their PDF format to a .txt File. The article's title, author(s) name and affiliation, reference list, tables and graphs were removed then the .txt files were uploaded to Sketch Engine under the corpus AEC. The corpus was then automatically tagged using English TreeTagger Part-of-speech (POS) tagset with Sketch Engine modifications. The corpus consists of files for each journal, each has its own meta-data which identifies the journal and its publication dates. The total size of the corpus was 1,204,791 tokens, 987,387 words.

Table 4-1: English Journals in AEC

	Journal	Number of Articles	Date	Number of Words
1	Applied Linguistics (Oxford University Press)	14	2013-2016	107,040
2	Studies in Language (John Benjamins Publishing Company)	12	2014-2016	100,958
3	Journal of Linguistics (Cambridge University Press)	14	2012-2016	122,391
4	TESOL Quarterly (Wiley-Blackwell on behalf of TESOL International Association)	15	2015-2016	103,050
5	Discourse & Society (Sage Publications)	15	2014-2016	104,523
6	Journal of Sociolinguistics (Wiley-Blackwell)	15	2014-2016	107,860
7	Language in Society (Cambridge University Press)	15	2014-2016	109,783
8	English Language & Linguistics (Cambridge University Press)	15	2014-2016	124,858
9	Discourse & communication (Sage Publications)	15	2014-2016	106,924
	Total	130		987,387

4.1.2 Academic Arabic Corpus (AAC)

A comparable Arabic corpus was compiled from a variety of journals in the field of Linguistics and Applied Linguistics. Those journals were accessed through the Saudi Digital Library (SDL). SDL is one of the largest Saudi websites that includes online reference books, online dissertations, and provides access to a large number of English and Arabic databases in a variety of fields and specialities. This library was created by the Saudi Ministry of Education and it is made available for all Saudi University students and employees. Through this library the researcher accessed DAR ALMANDUMAH database, which provides access to a number of journals in a variety of fields including Language and Applied linguistics. Thus, based on the availability, the researcher was able to access nine journals in the field of language and applied linguistics and select 10-15 articles from

each based on the same criteria used for the English journals. That is, the articles must be recent (published in the last 8 years), and a mix of quantitative and qualitative research papers. As with the AEC corpus, the articles tackle topics related to applied linguistics, language teaching, theoretical linguistics, and sociolinguistics. Table 4-2 presents the journals, the number of articles from each journal, their dates, and word size.

The same initial cleaning of the title, author's name and affiliation, graphs and tables took place. Then the articles were saved as .txt files, and uploaded to Sketch engine under the corpus acronym AAC. The corpus was then automatically tagged using Stanford Arabic POS tagset. The total size of the corpus was 1,154,020 tokens, 986,762 words.

Table 4-2: Arabic Journals in AAC

	Journal	**Number of Articles**	**Date**	**Number of Words**
1	Journal of Language Studies- University of Constantine 1 (Language Studies Laboratory-Algeria)	15	2010-2014	85,913
2	Al-Tareeb (Arabization) (The Arab Center for Arabization, translation, composition, and publishing-Syria)	14	2010-2015	90,428
3	Arabic Language (Supreme Council of the Arabic Language-Algeria)	10	2009-2010	80,588
4	The Jordanian Journal in Arabic language and literature (Mutah University- Deanship of the Academic Research- Jordan)	15	2012-2013	116,968
5	Language Sciences (Dar Gharib for Printing, Publishing, and distribution- Egypt)	15	2009-2011	189,241
6	Journal of the Saudi Scientific Society of Arabic Language (Imam Mohammed Bin Saud Islamic University- Saudi Arabia)	10	2009-2016	112,075
7	Journal of the Algerian Academy of Arabic Language (The Algerian Academy of Arabic Language- Algeria)	12	2009-2012	81,841

	Journal	Number of Articles	Date	Number of Words
8	Journal of the Faculty of Arabic language at Zagazig (Al-Azhar University- Egypt)	10	2009-2014	131,557
9	Journal of the College of Arabic Language (Omdurman Islamic University- Sudan)	15	2009-2014	98,151
	Total	116		986,762

4.1.3 Procedures

In response to the first research question of Study 2 (Which of the selected clause structures are frequently used in expert academic writing in both languages?) the analysis of the data should have started with identifying the most frequent clause structures using the Corpus Query Language (CQL) function available in Sketch Engine which enables users to input complex search queries for specific structures or collocations. However, the researcher encountered difficulties performing this step. Writing search queries for each of the selected clause structures resulted in a vast number of concordance lines which included examples of other clause structures and it was impossible to go through all of these lines and isolate the clause structures sufficiently. For example, using the query [tag="V.*"][tag="IN"], to search for the Prepositional Type 1 clause structure resulted in 29,608 concordance lines which included besides the prepositional type 1 verbs, such as *depend on* and *based on*, a number of other free combinations, such as *found in, come from* and *is in.* Furthermore, performing such queries in the Arabic corpus, the AAC, was more difficult as the tagger was insufficiently sophisticated.

Given that automatic disambiguation of the different clause structures is currently not possible within the technical remits of Corpus Linguistics, the analysis of the data started therefore by identifying the most frequent verbs in the two corpora and exploring how these verbs are used in the selected clause structures. With AEC, this step was conducted through the Word list function, then, the attribute "lempo" from the dropdown list of the search attribute. This attribute is a combination of lemma and part of speech. It specifies the output to a list of lemma forms, hyphen and a one-letter abbreviation indicating the part of speech of the lemma (e.g. be-v). At the filter option, the regular expression .*-v was used. This resulted in a list of lemmatised verbs ranked in order of

frequency. The 100 most frequent verbs were selected for further analysis in this study. That is because these verbs would represent the most frequently used verbs in this corpus of academic writing that would be of more benefit for learners to get familiar with as compared to less frequent verbs. Furthermore, for pedagogical purposes, it is more useful to focus on 100 verbs than exhausting the learner with information about all the verbs used in the corpus, given the fact that some of them are very rare and have very low occurrences. The total of the frequency of all the verbs is 144,371 and the sum of the frequency of the top 100 verbs is 94,832, which means that the sample of the study represents 66% of the total verb frequency in the AEC. The 100 most frequent verb lemmas, their raw and normalised frequencies are presented in Appendix 3a.

Modal verbs, e.g. *may, can, might,..*, were not included in this analysis. That was because modal auxiliary verbs could act only as auxiliary and were normally followed by a main lexical verb and not nouns. Thus, they were deemed not relevant for the purpose of the present study.

The same procedure was adopted to identify the most frequent verbs in the AAC. However, the lemmatisation of the Arabic verbs was conducted manually. This was because the Stanford Arabic POS tagset used to tag the Arabic data does not currently perform lemmatisation. This is possibly due to the syntactic ambiguity of Arabic associated with many factors including the devocalization of the conventional orthographic and the morphological richness of the language (Green & Manning, 2010). Accordingly, the researcher manually lemmatised the list of the most frequent verbs which was created using the CQL formula:

[tag="V.*"]

This search formula resulted in 124,962 concordance lines. Using the Node forms function, these verbs were ranked according to their frequency. The next step was to work on this list of verbs, lemmatise and rank them again. This step was conducted manually using Excel Worksheet. Verb forms that belong to the same root, trilateral being more common in Arabic or quadrilateral, were grouped under one lemma. Trilateral verb roots consist of a sequence of three consonants, for example the root *k-t-b (write)* is made of three main consonant letters, a number of prefixes and suffixes may be added to this trilateral root to derive many verbs that express different tenses, gender and number, such as *sa-yaktubu*-FUT.3SG.M *(he will write), taktubu*-PRES.3SG.M/F *(he/she is writing),*

yaktubuna-PRES.3PL.M *(they are writing)*. Quadrilateral verb roots consist of four basic consonants, such as *t-r-dʒ-m (translate)*, from which similarly many verbs are formed such as *sa-yutardʒim-u-* FUT.3SG.M *(he will translate), tardʒamat*-PAST.3SG.F *(she is translating), yu-tardʒimuna*-PRES.3PL.M *(they are translating).*

The 100 most frequent Arabic verbs were then translated into English using two Arabic-English dictionaries: Awde and Smith (2004) Arabic Dictionary: Arabic-English, English-Arabic and Wortabet and Porter (1995) Hippocrene standard dictionary: Arabic-English, English-Arabic. The final list of the 100 most frequent, lemmatised, transcribed and translated verbs ranked in order of frequency is presented in Appendix 3b. The sum of all verbs frequencies is 96,221 and the total of the frequencies of the top 100 verbs is 44,236. This means that the sample investigated represents 45% of the total of all verbs frequencies in the AAC.

Selecting the most 100 frequent verbs from experts' corpora reveals interesting information regarding the verb diversity in the two languages. While the 100 most frequent verbs account for more than half of the total of all verbs occurrences in the English academic corpus, the same number of verbs represents a smaller portion of the total of verbs in the Arabic academic corpus. This may indicate a higher verb diversity in the Arabic corpus.

The next step was to investigate the concordance lines of each verb in English and Arabic and identify the clause structures in which the verbs were most commonly used. For verbs whose frequency was over 1,000, a sample of 1,000 concordance lines was examined. In order to make the sample as representative as possible, it was created by including 111 concordance lines from each journal and 112 lines from the Journal "Discourse and Communication" for reasons of convenience as it is the last journal in the corpus. Selecting the concordance lines from each journal was easy using the metadata available on the left of each concordance line (see Figure 4-1). In the case of the Arabic corpus, the sample was selected from each word form of a lemma (see Figure 4-2).

The researcher investigated the concordance lines of each of the selected verbs in both corpora and classified the verbs according to the clause structures they were used in. In order to validate this classification, more than 30% of the English data was cross-checked by an English native speaker. Due to time and availability constraints, 10% of the Arabic data was checked by an Arabic native speaker.

Figure 4-1: Examples showing how metadata appear on Sketch Engine for the AEC

Figure 4-2: Example of how the sample is created for the verb /kana/ (be) in the AAC

Verb	Transcription	Meaning	Frequency	Total frequency	sample
كان	/kana/	be	3,509	8376	500
يكون	/yaku:nu/		1,917		200
كانت	/kanʌt/		1,107		100
تكون	/taku:nu/		1,052		100
يكن	/yaku:nu/		318		20
كانوا	/kanu:/		193		20
تكن	/takun/		133		20
كنت	/kuntu/		96		20
كنتم	/Kuntum/		51		20

4.1.4 Data processing and statistical analysis

Given that the sizes of the two corpora were not exactly equivalent, AEC (1,204,791 tokens, 986,762 words) and AAC (1,154,020 tokens, 987,387 words), one more step was taken when comparing frequencies of occurrence in the two corpora which was to calculate the normalised frequency (norm freq). A good base of normalisation for the given corpora sizes would be one million, especially that frequency per million was the most commonly encountered in the literature (McEnery & Hardie, 2012: 49-50). However, because of the comparison between two other corpora of novice writers in the following chapter which sizes were 51,761 tokens (44,090 words) for the native speaker corpus and 46,668 tokens (40,989 words) for the advance Saudi learner corpus, a common base of normalisation was

used which is 100,000. The normalised frequency was calculated using the following formula:

Raw frequency ÷ corpus size ×100,000

The corpus size value used in the above formula was the number of words in each corpus, i.e. 986,762 words for AEC and 987,387 words for AAC, not the number of tokens. That is because in sketch engine the number of tokens included word forms as well as punctuation while our focus is on word forms only.

In order to determine the level of contingency between the verb and the clause structure in which it was used, a further step was taken, which was to normalise the number of occurrences of each verb in a particular clause structure over the total frequency of occurrence of that verb per 100.

Raw frequency of the clause structure ÷ total frequency of the verb ×100

This measure, known as 'Measure of Faithfulness' by (Römer et al., 2015), reflects the proportion of the total tokens frequency of the verb that appeared in a clause structure. This procedure provides a clearer picture of the percentage of use of this verb in a given clause structure and the level of 'faithfulness' of the verb to that clause structure. For example, if the verb *become* has a measure of faithfulness of 98 to the copular clause structure while the verb *get* has the measure of faithfulness of 30 to the same clause structure, this indicates that the verb *become* is more 'faithful' to the copular clause structure than the verb *get*, which seems more versatile and may be more faithful to other clause structures. This measure is particularly useful for language learners and educators because, as explained by Römer et al. (2015: 56), clause structures with more faithful members are more easily learned. Knowing which verbs are particularly faithful to which clause structure or structures could provide a much more comprehensive understanding of the usage of verbs combining both syntactic and semantic (meaning) features. The phraseological approach to lexical items in academic writing, e.g. verb collocations, produces mostly lists with frequent items (e.g. Brezina & Gablasova, 2017), sometimes supplemented with their most frequent senses. This is useful in that it shows which items are likely to occur in a particular register or genre and which meanings are the most prevalent. Yet, based on just frequency lists learners would not know how to use the items correctly and appropriately in texts (Stein, 2017). Learners' knowledge of the faithfulness

of a verb to a particular clause structure or structures can help learners in understanding the appropriate use of the verb in the most preferred context.

When comparing the use of clause structures in two corpora, the Log-likelihood test (LL) of statistical significance was used. The LL was also used in study 1 because it is the preferred test for significance for corpus studies as it does not assume that data is normally distributed. The LL value is significant at the level of $*p<0.05$ when it is greater than 3.84, at the level of $**p<0.01$ when it is greater than 6.63, at the level of $***p<0.001$ when it is greater than 10.83, and at the level of $****p<0.0001$ when it is greater than 15.13, when the degree of freedom is df =1, as we were comparing two corpora. (http://ucrel.lancs.ac.uk/llwizard.html).

In order to report the effect size, Bayes Factor (BIC) is calculated using the same Lancaster LL wizard. The BIC value of 0-2 is not worth more than a bare mention, 2-6 shows positive evidence against H0, the null hypothesis which states that there is no statistically significant difference, 6-10 shows strong evidence against H0, >10 indicates very strong evidence against H0.

It was important to group the verbs into semantic classes because this helped in highlighting the overarching semantic classes used by expert writers and in the comparison between experts' and novice writers use of these semantic classes, to be conducted in study 3. Therefore, for this purpose, Hinkel (2004: 178-200) classification of academic verbs into five classes was adopted. The five classes included activity verbs (*make, create, use*), reporting verbs (*suggest, say, discuss*), mental verbs (*think, consider, see*), linking verbs (*become, appear, seem*) and logico-semantic relationship verbs (*include, reflect, result*). Hinkel's classification was particularly preferred because it focuses on the functions of verbs commonly used in academic writing and attempts to give a transparent clear label for each semantic class.

For the semantic frames analysis, the researcher mainly referred to the FrameNet database (https://framenet.icsi.berkeley.edu/fndrupal). FrameNet is a lexical database based on the theory of frame semantics that facilitates the retrieval of the semantic frame(s) of verbs based on their senses. It provides the sense and the semantic frame of the verb as well as the semantic roles associated with that frame. It also indicates which of these roles are core roles and which are peripheral or non-core. For example, the phrasal verb *take over* activates the semantic frame of 'change of leadership', this semantic frame involves three

core roles: old leader, new leader, and a function/role. Identifying these roles in the concordance lines of *take over* results in a more comprehensive understanding of the collocations of that phrasal verb.

Unfortunately, work on the FrameNet project is not complete, and many senses of verbs are not included in the database. In such cases, instead of starting with the semantic frame, the researcher took the verb as the starting point and tried to identify the lexical items that filled in the syntactic positions of the subject and the object and the semantic roles that these elements represent. Dictionaries, such as Cambridge Dictionary (http://www.cambridge.org), The Oxford English Dictionary (http://www.oed.com), and Merriam-Webster (https://www.merriam-webster.com) were also referred to.

It is important to mention that when the passive voice is used, the researcher converted the passive clause structure into its active form, so that the original subject was identified, if mentioned in the regular passive form in a *by*-phrase. If the original subject was not mentioned because the stative passive was used, the researcher reported the subject as unknown. The subject could also be unknown because the verb is used in a nonfinite clause, such as the infinitive *to*, or the *-ing* participle.

A special section was assigned for the lexico-grammatical analysis of phrasal type 2, phrasal prepositional type 1,2, and 3, and prepositional type 3 clause structures. The analysis aimed to investigate the use of verbs in academic writing as compared to general use. General language use is reported in the PHaVE list which is based on the BNC and the COCA (Garnier and Schmitt, 2015). Therefore, the researcher compared the senses mentioned in that list for each phrasal verbal to the senses reflected in the concordance lines under investigation. Since the PHaVE list is mainly concerned with phrasal verbs, phrasal prepositional and prepositional type 3 verbs are not listed. In such cases, the above-mentioned dictionaries were used as a reference for general use.

4.2 Methodology for Study 3

In study 3, the characteristics of learners' authentic written production were compared with those of native English students. To achieve this aim, exam papers were chosen as the source of data in order to ensure that the writing is learners' genuine work not influenced by the use of additional resources such as dictionaries or webpages, or teachers' feedback. The exam papers were photocopied, anonymised and number coded.

They were then transcribed and uploaded to Sketch Engine. The corpora were then automatically tagged using the same tagset used for the AEC, namely the English TreeTagger Part-of-speech (POS) tagset with Sketch Engine modifications. In the process of transcribing the essays, care was taken not to correct or change any of the participants' production so as to keep the raw data (Granger, 2002). Quotes in the essays were deleted, so that quoted material would not affect the participants' production. More details about the two corpora is presented in the following section.

The exam papers were collected during the academic year of 2015-2016. Out of the thirteen genre families identified in Nesi and Gardner (2012), students' answers included in this study may be classified under three main genre families, explanations, essays, and critiques. In the explanation genre students demonstrate their understanding of a certain concept or approach in applied linguistics, such as giving an explanation of the audiolingual method or the direct method. In critiques, students give an evaluation of a certain phenomenon, such as linguistic variation. In essays, students make connections and build their own argument to topics such as who leads language change, men or women.

4.2.1 Novice Native Corpus (NNC)

The novice native (NNC) corpus was compiled from native speaker students' exam essays for different applied linguistics modules taken at the Department of English Language and Applied Linguistics at the University of Reading in the year 2016. Forty participants were drawn from four different modules, *Literacy*, *English Grammar and Lexis, English in the World,* and *Sociolinguistics.* The length of the essays ranged from 500-2000 words. To control for the proficiency level of the essays included in the corpus, only essays that scored 60% or above were included. The researcher aimed for thirty samples from each module, however, this aim was not possible due to the limited number of native students enrolled for each module. Therefore, with the priority of compiling a comparable corpus size to that of advanced Saudi learners, following Nesi and Gardner (2012), two well-developed essays of some exam papers were used as separate texts. The following table (Table 4-3) illustrates the number of exam papers, the number of texts and the total word count for each module. The total size of the corpus was 51,761 tokens, 44,090 words.

Table 4-3: The NNC

Module	Exam papers	Texts	Words
English Grammar and Lexis	10	14	11,748
English in the world	10	13	11,085
Literacy	9	12	9,951
Sociolinguistics	14	16	11,306
Total	43	55	44,090

4.2.2 Novice Saudi Corpus (NSC)

A comparable corpus was compiled from exam papers written by novice advanced Saudi learners at the department of English Language and Literature at the Bachelor degree at Imam Abdulrahman Bin Faisal University in the year 2016. Samples from four different modules were gathered. These modules were *Applied Linguistics, Psycholinguistics, Sociolinguistics,* and *Semantics.*

The same criteria for the selection of the exam papers for the NNC were applied. The length of the essays was 500-2000 words and only essays that scored 60% or above were included in the corpus. Table 4-4 presents the number of exam papers, the number of texts and the total word count for each module. The total size of the corpus was 46,668 tokens, 40,989 words.

Table 4-4: The NSC

Module	Exam papers	Texts	Words
Applied linguistics	20	20	11,511
Psycholinguistics	24	24	10,527
Semantics	30	30	7,343
Sociolinguistics	30	30	11,608
Total	104	104	40,989

4.2.3 Ethical procedures

Before collecting samples from students, both native and Saudi, Ethics committee approval was obtained (see Appendix 5a for the NNC and 5b NSC). Students were provided with an information sheet that explained the project and its arrangements (see appendix 5c) then, they were requested to give their permission to the researcher to use their written exam papers by signing the consent form provided in appendices 5d for the NNC and 5e

for the NSC. To ensure confidentiality, students' personal data (such as names and contact details) were only available to the researcher and supervisors and exam papers were photocopied, anonymised and given number codes instead of the students' names, such as 1-1, 1-2, 1-3 and so on.

4.2.4 Procedures

Similar to Study 2 and due to the same difficulties of performing search queries for each clause structure, the analysis of the data of Study 3 started by identifying the 100 most frequent verbs in the two corpora and exploring how these verbs are used in the selected clause structures. The most frequent 100 verbs used by novice native learners in the NNC were identified using a similar procedure to that used in study 2 to extract the verbs from the AEC. Using the word list function in Sketch Engine, then, the attribute lempo from the dropdown list of the search attribute, and the regular expression .*-v at the filter option resulted in a list of lemmatised verbs ranked in order of frequency. The total frequency of all verbs in the NNC is 7,035 and the sum of the frequency of the 100 most frequent verbs is 4,563, which means that the sample represent 64% of the total verbs frequency.

Following the same procedure, the 100 most frequent verbs used by novice advanced Saudi learners in the NSC were identified. The sum of these 100 verbs is 5,041 compared to a total of 6,640 occurrences of all verbs in the NSC. This means that the sample investigated represent 76% of the total verb frequency.

It is worth mentioning here that this step of selecting the 100 most frequent verbs from the English corpora investigated in this thesis, the AEC, the NNC and the NSC, results in a sample that consistently represent more than 60% of the total occurrences of all verbs. This indicates that verb diversity is similar in all English corpora regardless of the level of expertise of the writers. It is also noted that in the novice Saudi learners' corpora, the NSC, the 100 most frequent verbs represent a higher percentage than that of the AEC and the NNC, which indicates lower verb diversity in the NSC as the 100 most frequent verbs account for more than three-quarters of the total of all verbs used in the corpus.

Full lists of the hundred most frequent verbs in the NNC and the NSC along with their raw and normalised frequencies are presented in Appendix 6a for the NNC and 6b for the NSC.

The next step was to investigate the concordance lines of each verb in the two corpora in order to identify the clause structures in which the verbs were used. For the verb *be*, because its frequency in the two corpora is over 1000, only a sample of 1000 concordance lines was examined. This sample was created by including 250 lines from each module. In order to validate this classification, data was cross-checked by an English native speaker.

A number of concordances in the NSC were viewed as errors by the researcher. Thus, these concordances were extracted and listed in a table to be checked by two native speakers in order to confirm their identification as errors. It was important to consult two raters rather than one because the two raters' agreement ensures that their judgement is not affected by any factor related to one of the raters, such as their background or personal experience (Gwet, 2014:4). The two raters' agreement maximises the inter-rater reliability (IRR). The two raters, Chris Ryder and Sophie Payne, who are PhD students at the departments of English language and Applied linguistics and Modern Languages and European studies, respectively, kindly agreed to read the concordance lines and judge whether the use of the verb in each concordance was an error or not (Yes/No) and provide reasons why a concordance was considered unacceptable or provide correction. The IRR was calculated using the Cohen's Kappa statistical measure. This measure was used because it is a preferred measure to assess IRR for fully-crossed nominal data, such as the data at hand (Hallgren, 2012). The IRR was found to be Kappa= 0.927, which is considered a near perfect agreement (Landis & Koch, 1977). The two raters disagreed on three items only and in all three the first rater was not sure whether it is an error or not (Yes and No), thus the item was considered an error. A total of 90 items were listed to be checked by the raters. The raters agreed on the acceptability of 7 items only and the rest of the items were all considered errors and were excluded from the data and handled separately. Appendix 8 presents a table of all the checked items along with the raters' judgement.

4.2.5 Data processing and statistical analysis

In accordance with the procedure used in study 2, given the fact that the sizes of the two corpora are not exactly equivalent, NNC (44,090 words) and NSC (40,989 words), the first step for processing the data was to calculate the normalised frequencies (norm/freq). The same base of normalisation used in study 2 which was 100,000, was used to calculate

normalised frequency for all raw frequencies in Study 3. Using the same base of normalisation for Studies 2 and 3 facilitates the comparison of the results of the two studies.

For statistical significance, as in study 2, the Log-likelihood test (LL) of statistical significance and the BIC values of effect size were calculated using the Lancaster LL wizard used (http://ucrel.lancs.ac.uk/llwizard.html). A further step was taken, as was the case with the expert corpora in study 2, the measure of "Faithfulness" (Römer et al., 2015) was calculated. This measure helps in analysing the data in response to Q2 which entails the identification of the verbs associated with each clause structure.

For the classification of the semantic classes of verbs, as in study 2, Hinkel (2004: 178-200) classification of academic verbs was used. For the semantic analysis, the researcher followed the same procedure explained in section 4.1.4 for the semantic analysis of Study 2. That is FrameNet database was mainly used. In the cases where the verb is not covered in the database, dictionaries were consulted.

Chapter 5: A contrastive analysis of verb complementation clause structures and verb-noun collocations in English and Arabic Expert writers' corpora

5.1 Introduction

This chapter presents an analysis of verb-noun collocations within the frame of verb complementation clause structures in expert academic writing in two languages, English and Arabic. The analysis explores the occurrences and use of verb-noun collocations adopting a novel analytical approach which employs Quirkian clause structures (Quirk et al,1985) and draws on Frame Semantics (Fillmore, 1982). Such analysis presents a major shift in general research concerned with collocations from the previous focus on mere co-occurrence of lexical items to an in-depth analysis of the use of specific lexical items in specific syntactic clause structures. Grounded in the understanding that forms and meanings are essentially tied together (Sinclair, 1991; Hunston & Francis, 2000; Goldberg, 1995), this analysis highlights the lexico-grammatical features of verb complementation clause structures and VN collocations in expert academic writing and sets out benchmarks for further analysis of learner data.

5.2 Analysis based on total frequencies

The concordance lines of each of the 100 most frequent verbs in Arabic and English are classified according to the selected 15 clause structures. These selected clause structures are presented in Table 5-1, along with their frequencies, raw and normalised, in both corpora, the LL values and the BIC values. Normalised frequencies are provided between brackets. The bold font is used where there is a significant difference between the two groups to highlight which group used more tokens/ types. If a clause structure does not exist in Arabic, such as phrasal verbs, the clause structure is represented by an X. However, if the clause structure is possible in Arabic, but is not found in the data at hand, a value of (0) is given to that clause structure, as in the case of the prepositional type 2a. A detailed presentation of verbs classified under each clause structure, their raw frequencies, normalised frequencies, and their measure of faithfulness to the clause structures in both corpora is presented in Appendix 4. Attached to this thesis is a CD that includes table reports of the use of each verb in the two corpora, checklists of the clause structures and

the verbs used in each clause structure along with their raw frequency, as well as all the coded concordance lines.

Table 5-1: Selected clause structures and their frequencies in AEC and AAC

		Examples	Types		LL Value	BIC	Tokens		LL value	BIC
			AEC	AAC			AEC	AAC		
1	**Copular/Linking** SVC (_VP NP/ ADJ)	*She is a teacher The girl seemed restless*	11 (1)	9 (1)	0.20	-14.30	**2415 (245)**	1786 (181)	94.14 ****	79.64
2	**Simple Transitive** SVOd (_VP NP)	*Tom caught the ball*	86 (9)	73 (7)	1.06	-13.44	**26998 (2734)**	11569 (1172)	6338.91 ****	6324.41
3	**Phrasal Type 2** SVOd (_VP AdvPart NP)	*Sam picked up the pen Sam picked the pen up*	41 (4)	X	X	X	252 (26)	X	X	X
4	**Prepositional Type 1** SVOp (_VP PP)	*John looked at his watch*	**75 (8)**	46 (5)	7.00 **	-7.50	3476 (352)	**4282 (434)**	86.62 ****	72.12
5	**Phrasal Prepositional Type 1** SVOp (_VP AdvPart PP)	*I look forward to your party*	25 (3)	X	X		78 (8)	X	X	X
6	**Complex Copular** SVOC (_VP NP NP/ADJ)	*He considered his uncle a genius/ He found the book relevant*	**45 (5)**	9 (1)	26.18 ****	11.68	**3080 (312)**	2655 (269)	31.26 ****	16.76
7	**Complex Transitive** SVOA (_VP NP PP)	*She put her coat in the hall*	4 (0)	2 (0)	0.68	-13.82	**138 (14)**	66 (7)	25.92 ****	11.43
8	**Ditransitive/ Double object** (dative, benefactive and depriving clause structures) SVOiOd (_VP NP NP)	*He lent Sam his bike*	**9 (1)**	2 (0)	7.36 **	-7.14	**299 (30)**	13 (1)	324.26 ****	309.77
9	**Prepositional**	*He lent his bike to Sam*	**6 (1)**	0	8.31 **	-6.18	**116 (12)**	0	160.74 ****	146.24

		Examples	Types		LL Value	BIC	Tokens		LL value	BIC
			AEC	AAC			AEC	AAC		
	Type 2a (alternating) SVOdOp (_VP NP PP)									
10	**Prepositional Type 2b** (non alternating) SVOdOp (_VP NP PP)	*He donated £10 to charity*	**23 (2)**	8 (1)	7.56 **	-6.93	177 (18)	172 (17)	0.07	-14.43
11	**Prepositional Type 3** SVOdOp (_VP NP PP)	*I caught sight of him*	**29 (3)**	2 (0)	28.13 ****	13.63	**517 (52)**	16 (2)	594.87 ****	580.38
12	**Prepositional Type 4a** (animate indirect object) SVOiOp (_VP NP PP)	*They told me about your success*	8 (1)	4 (0)	1.36	-13.14	**124 (13)**	73 (7)	13.32 ***	-1.17
13	**Prepositional Type 4b** (inanimate indirect object) SVOiOp (_VP NP PP)	*They based the findings on fact*	**19 (2)**	4 (0)	10.62 **	-3.87	**1488 (151)**	53 (5)	1674.01 ****	1659.51
14	**Phrasal Prepositional Type 2** SVOdOp (_VP NP AdvPart PP)	*They put it down to chance*	6 (1)	X	X	X	23 (2)	X	X	X
15	**Phrasal Prepositional Type 3** SVOiOp (_VP NP AdvPart PP)	*They let me in on the deal*	5 (1)	X	X	X	10 (1)	X	X	X
Total of matched clause structures			**315 (33)**	158 (15)	53.01 ****	38.52	**38818 (3932)**	20685 (2095)	5603.27 ****	5588.77

The distribution of the most frequent 100 verbs in AEC and AAC over the clause structures, presented in Table 5-1 and Figures 5-1 and 5-2, reveals a number of syntactic

features of academic writing in the field of applied linguistics in English and Arabic and points to several similarities and differences between the two languages.

Out of the 100 most frequently used verbs, English writers use 86 verbs in the transitive clause structure and 73 verbs are used by Arabic writers in this type of structure. This may point to the fact that this type of clause structures is the most frequently used clause structure in the sample of academic writing in both languages. English writers use the transitive clause structure more frequently than Arabic writers as they use more types of verbs and significantly more tokens therein.

Another large group of the 100 most frequent verbs occurs in the prepositional type 1 clause structure in the AEC (75 types) and the AAC (46 types) suggesting that this clause structure is the second most frequently used clause structure by both groups of expert writers. Arabic writers use significantly more tokens in the prepositional type 1 clause structure than English writers (434 to 352, respectively) with an LL value of 84.62**** and a greater than 10 BIC value indicating a very strong statistically significant difference. However, English writers use significantly more types in this clause structure than Arabic writers (8 to 5 with LL 7.00**) suggesting that this clause structure in Arabic academic writing operates with a smaller range of verbs that are more frequently re-used. Further analysis of this clause structure reveals that while in Arabic a verb is used with either one or two prepositions resulting in one to two types, verbs in English tend to occur with a greater variety of prepositions which results in a larger number of types. For example, the verb *come* is used with 7 prepositions: *to* (33), *into* (24), *with* (18), *in* (4), *across* (2), *by* (2), *under* (1), and the verb *go* is used with 8 prepositions: *through* (17), *for* (6), *into* (6), *with* (4), *against* (3), *over* (2), *to* (1), *about* (1).

In the complex copular clause structure, English writers use significantly more types and tokens than Arabic writers (5 to 1 normalised types and 312 to 269 normalised tokens, respectively). This significant difference, which is also strongly supported by the greater than 10 BIC value, indicates the importance of the complex copular clause structure in English academic writing in the field of applied linguistics. The use of this clause structure in the sample of Arabic academic writing is restricted to a limited number of verbs, such as */dʒaʕala/ (make/ create), /juʕaˈdu/ (consider/ is no longer/ count),* and */saˈma/ (call).* However, a larger number of verbs are used in this same clause structure in the

sample of English academic writing, for example, *make*, *consider*, *call*, *define*, *find* and *understand.*

Figure 5-1: Normalised frequencies of types in the selected clause structures in AEC & AAC

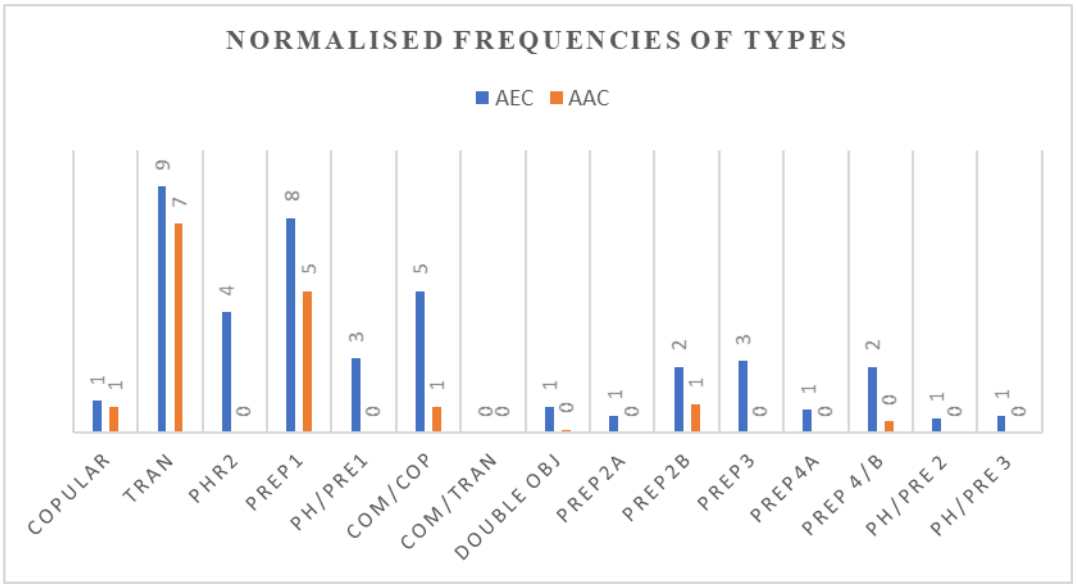
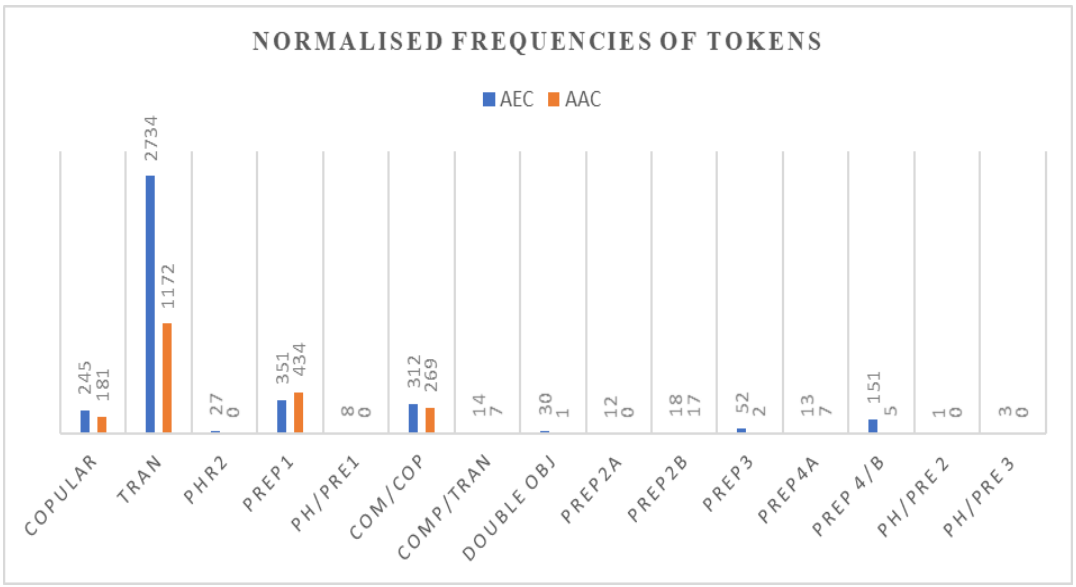


Figure 5-2: Normalised frequencies of tokens in the selected clause structures in AEC & AAC



The copular clause structure is more frequently used by English expert writers. More verbs are used in this clause structure in the AEC (11) than in the AAC (9). And although the normalised frequencies show no difference in the types, both values

normalised to 1, the English writers used significantly more normalised tokens (245 compared to 181).

The frequency of types and tokens in the double object clause structure and its alternation, prepositional type 2a, is significantly higher in the AEC than in the AAC. While the BIC values reduce the importance of the significant difference between the frequency of verb types in these two clause structures, the greater than 10 BIC values in relation to verb tokens reflect the importance of the significant difference between the tokens of verbs used in these two clause structures. These two clause structures are available in Arabic, but they seem to be more common in English academic writing. One explanation for this difference is that the verb */ʔaʕtˁa/(give)* which is the prototype verb of this structure, in English as well as in Arabic, is not one of the most frequent verbs in the AAC; it has the normalised frequency of 2 (24 raw freq).

In general, the distribution of tokens and types, in Figures 5-1 and 5-2, shows that English writers rely on a variety of clause structures, whereas Arabic writers rely on a limited set of clause structures. In the AAC, less syntactical diversity is observed. Verb types are mostly used in the transitive clause structure and prepositional type 1 and fewer types are used in other clause structures. This can be attributed to the fact that some of these clause structures, such as phrasal and phrasal prepositional clause structures, do not exist in Arabic.

Furthermore, the distribution of the most frequent verbs in the AEC highlights the importance of phrasal verbs in Academic English writing in Applied Linguistics, which emerged as an unexpected result. Contrary to the claims of Biber, Conrad, and Leech (2002:127), Biber et al. (1999), and Liu (2011) that phrasal verbs are rare in academic writing, the analysis above shows that 41 types are used in phrasal type 2 clause structure, 25 in phrasal-prepositional type 1, and 11 in phrasal prepositional type 2 and type 3. These form a total of more than 75 types of phrasal and phrasal prepositional verbs which accounts for nearly 20% of the total of verb types identified for the 100 most frequent verbs in the AEC which is 392. The total of tokens of phrasal verbs identified for the 100 most frequent verbs in the AEC is more than 300 tokens which combines the sum of 252 tokens in phrasal type 1, 78 tokens in phrasal prepositional type 1 and 33 tokens in phrasal prepositional type 2 and 3. Compared to the total of all tokens of verbs in the other clause structures investigated which is more than 30,000 tokens, phrasal verbs represent an approximate ratio

of 1 phrasal verb for every 100 tokens. This result has not been anticipated and came across as rather surprising. Generally, phrasal verbs are more common in conversation and fiction registers and less so in academic writing (Biber et al. 2002). In fact, writers - expert or novice - are discouraged from using phrasal verbs which are seen as a feature of informal registers (Siyanova & Schmitt, 2007). Yet, the data under investigation is based on a recently published work and this might suggest that the use of phrasal verbs is becoming more common in academic writing in the field of applied linguistics.

The study explores the frequency of occurrence of the top 100 verbs in a variety of clause structures in the AEC and the AAC. Although the starting points are the same, 100 verbs for each corpus, the comparison of the total frequency of types in the matched clause structures, excluding phrasal and phrasal prepositional clause structures that do not exist in Arabic, shows a statistically significant difference (LL 53.01****). English writers use almost double the number of types used by Arabic writers (315 to 158). Token frequency also reveals a statistically significant difference (LL 5603.27****) which may be a consequence of the difference between the types. One reason for this difference in the frequency of types may be that some clause structures in English allow the use of a larger number of verbs whereas these clause structures are limited to a certain set of verbs in Arabic. For example, in the complex copular clause structure, only 9 verbs are used in the AAC compared to 45 in the AEC.

Another reason might be related to the use of prepositions in prepositional clause structures and the fact that in English, verbs tend to combine with more than one preposition. In the prepositional type 1 clause structure, 33 verbs are used in both the AEC and the AAC but the total of types in this clause structure in the AEC is 75 compared to a total of 46 types in the AAC. An example that may demonstrate this difference in the distribution of prepositions is the verb *go* and its Arabic equivalent */ðahaba/(go)*. While *go* is used in the AEC with 8 prepositions resulting in 8 types for the prepositional type 1 clause structure, (*go through* (17), *go for* (6), *go into* (6), *go with* (4), *go against* (3), *go over* (2), *go to* (1), *go about* (1)), the verb */ðahaba/(go)* is used with one preposition *ila (to)* resulting in one type.

A third reason for the difference in the total frequency of types may be attributed to the intransitive use of a number of verbs in the AAC. More than 10 verbs of the top 100 verbs investigated in this study are used in the intransitive clause structure in more than

80% of their total occurrences. Since investigating the intransitive clause structure is beyond the scope of this study, these occurrences were excluded.

The significant difference between the total types and tokens may also be attributed to the exclusion of more than 1500 concordance lines during the classification process because the verb form was used as either a common noun or a proper noun, for example, the verb */zada/(increase)* (which had a total raw frequency of 144) occurred as the proper noun (a name of a person) */jazi:d/* in 60 occurrences, the verb */jaʕi:ʃu/ (live)* (118) was also used as a proper noun (a name of a person) in more than 90 occurrences.

5.3 Analysis based on the measure of faithfulness

The measure of faithfulness is an important indicator of the relationship between the verb and the clause structure in which the verb is used. If a verb has a high measure of faithfulness to a given clause structure, this means it is mostly used in this clause structure rather than other clause structures. For example, if the measure of faithfulness for a verb in a particular clause structure is 100%, this means that the verb occurs only in this clause structure. As explained in Section 4.1.4, information on the level of contingency between the verb and the clause structure is useful for language learners and teachers because it provides learners with a better understanding of the use of the verb in the most appropriate and preferred clause structure.

This analysis focuses on clause structures that have more faithful members, specifically, verbs that have the measure of faithfulness of 90% or more. By calculating the mean of the measures of faithfulness of each verb in each clause structure, which is presented in Appendix 4, it is observed that the copular and the transitive clause structures have higher means of faithfulness than other clause structures in both corpora, see Table 5-2, meaning that there is a range of verbs that mostly occur in these structures and not others. It is also observed that the complex copular and prepositional type 1 clause structures have higher means of faithfulness in both corpora than other clause structures and the mean of faithfulness of these two clause structures is higher in the AAC than in the AEC. That is why the rest of this section further examines these four clause structures.

Table 5-2: Means of the measure of faithfulness of the verbs to the 15 clause structures

		Mean of Faithfulness	
		AEC	**AAC**
1	**Copular/Linking** SVC (_VP NP/ ADJ)	41	50
2	**Simple Transitive** SVOd (_VP NP)	65	59
3	**Phrasal Type 2** SVOd (_VP AdvPart NP)	1	X
4	**Prepositional Type 1** SVOp (_VP PP)	11	36
5	**Phrasal Prepositional Type 1** SVOp (_VP AdvPart PP)	1	X
6	**Complex Copular** SVOC (_VP NP NP/ADJ)	15	59
7	**Complex Transitive** SVOA (_VP NP PP)	5	12
8	**Ditransitive/Double object** (dative, benefactive and depriving clause structures) SVOiOd (_VP NP NP)	7	4
9	**Prepositional Type 2a** (alternating) SVOdOp (_VP NP PP)	3	0
10	**Prepositional Type 2b** (non-alternating) SVOdOp (_VP NP PP)	2	8
11	**Prepositional Type 3** SVOdOp (_VP NP PP)	3	3
12	**Prepositional Type 4a** (animate indirect object) SVOiOp (_VP NP PP)	3	14
13	**Prepositional Type 4b** (inanimate indirect object) SVOiOp (_VP NP PP)	18	8
14	**Phrasal Prepositional Type 2** SVOdOp (_VP NP AdvPart PP)	1	0
15	**Phrasal Prepositional Type 3** SVOiOp (_VP NP AdvPart PP)	1	0
The Mean of Means		11.80	19.46

5.3.1 Copular

In the AEC, two verbs are more than 90% faithful to the copular clause structure, *become* (98%) and *seem* (97%), see Appendix 4. These two verbs are also ranked as the most frequently used verbs in this clause structure. The high frequency of the resulting verb *become* in the AEC, confirms with Biber et al.'s (1999: 443) observation that the verb *become* is extremely common in academic writing. Furthermore, according to Biber et al.

(1999:437), *become* is one of the copular verbs that most commonly take an adjectival complement. The analysis of its use in the AAC confirms Biber et al's claim, in more than 60% of its total occurrences in the copular clause structure, the verb *become* is taking an adjectival complement (395 occurrences) compared to 190 occurrences with a nominal complement, as in Example 1.

Example 1: *become* in AEC

become +adjective in AEC

studies indicate that university discourse has	**become**	more promotional and more similar to business
use of [BE Ving] from being uncommon to	**becoming**	predominant, which led to the construction

become +noun in AEC

policies (Tomlinson, 2005). Neoliberalism has	**become**	a key political ideology and the role of
studies. Indeed for many theorists it has	**become**	an essential component of the knowledge-based

The verb *seem* refers to the likelihood or strong possibility (Hinkel, 2004: 193). In academic writing, this verb performs the function of a hedge (Hyland, 1998). Its use in the AEC confirms this observation as it is used in almost 70% of its total occurrences followed by the *to-infinitive,* as in Example 2.

Example 2: *seem* in AEC

Another advantage of this kind of analysis	**seems**	to come from coordination and extraction
effect of noun type Another factor that	**seems**	to show a correlation with the choice of

In the AAC, the two resulting verbs, */sˁa:ra/* and */ʔsˁbaha/* with the meaning of *become*, are the most faithful verbs to the copular clause structure. Unlike English, the copular verbs in the AAC take nominal complement more frequently than the adjectival complement. This suggests that Arab advanced learners might not underuse this clause structure overall but might be less likely to use adjectival complements in this clause structure.

5.3.2 Transitive

The transitive clause structure has a high mean of faithfulness in the two corpora but the mean of faithfulness in the AEC is higher than that in the AAC (65 to 59, respectively). Twenty-eight verbs are found to be strongly faithful (90% and more) to the

transitive clause structure in the AEC and twenty-six verbs in the AAC. These verbs are classified into four semantic classes, see Table 5-3.

Table 5-3: Semantic classes of the most faithful verbs to the transitive clause structure in the AEC and the AAC

Semantic Class	AEC	AAC
Activity verbs	*demonstrate, produce, illustrate, reveal, show, establish, create, provide,*	*/istaʕmala/ (use), /ħamala/ (carry), /ħaðafat/(omit), /istaxdama/ (use), /ħawala/(try), /yuʃakılu/ (form), /ʕaradˁa/ (show), /tabaʕa/ (follow)/ /raħɜma/(have mercy on), /yantˁiqu/ (utter)*
Reporting verbs	*suggest, discuss, express, note, explain, address, claim, indicate,*	*/ðakara/(mention), /qala/(say), /tanawala/ (deal with), /naqala/ (report)*
Mental verbs	*examine, tend, explore, need, observe, determine, reflect*	*/ʔarada/(want), /qarʔa/ (read), /yufi:du/(benefit), /yadrusu/ (study), /qasˁada/ (intend/ aim), /yaʕni:/ (mean/ care/ pay attention), /χalafa/ (disagree), /samiʕa/ (hear), /yafhamu/ (understand),*
Logical semantic-relationship verbs	*involve, contain, include, support, represent,*	*/yaqtadˁi:/(require), /yatatˁalabu/(demand), /yatadˁamnu/ (include), /yuqabilu/ (confront),*

In academic writing, reporting verbs are particularly important because they are used in paraphrasing and citing information. Both processes are essential to show the writers knowledge and familiarity with external sources (Hinkel, 2004:187). Logical-semantic relationship verbs are also important for academic writing and are used to express advanced lexical functions of establishing and/or explaining the relationship between two constructions of knowledge (Hinkel, 2004: 195). Given the importance of reporting verbs and logical-semantic relationship verbs in academic writing, the use of verbs that belong to these two semantic classes is further scrutinised.

In the AEC, a larger number of reporting verbs are faithful to the transitive clause structure than in the AAC. These verbs include the verbs *suggest, note, claim* and *indicate* which are mainly complemented with a finite *that-clause.* The verb *explain* is another

reporting verb used in the AEC and it is mostly followed by a finite *wh-clause*. It collocates with *why* and *how* with a logDice value of (11.263 and 9.101, respectively) meaning that the verb *explain* is mainly used to provide the reason or the process for a given subject.

The verbs *discuss, express* and *address* are mostly followed by a noun phrase. *Discuss* is associated with the noun *section* with a logDice value of (9.560). The noun *section* is used either in the subject position or as part of a following prepositional phrase to indicate location, as in Example 3:

Example 3: *discuss+ section* in AEC

for the present analysis. Then, ***Section*** 3.2	**discusses**	tense and clausal grounding in CG. 3.1
The article is organized as follows. ***Section*** 2	**discusses**	key grammatical properties of the
texts. Possible explanations for this will be	**discussed**	in ***Section*** 6.2. Syntactic generalization In
in Hebrew. Data from other languages is	**discussed**	in ***Section*** 2.2 to illustrate some important

The verb *discuss* and the verb *express* both collocate with personal pronouns, e.g. *we* and *they*, in the subject position, as in Examples 4 and 5 suggesting that the process of discussing and expressing is also done in the first or third person.

Example 4: *discuss + we* in AEC

than it is in Russian. Therefore, ***we*** do not	**discuss**	such data in the remainder of the paper (for a
(e.g. Stump 1997). In Section 3, ***we*** have first	**discussed**	the literature on morph ordering that attempts

Example 5: *express+ they* in AEC

to expressing exclamations,although ***they*** can	**express**	various indirect speech acts based on these
convey subjective meaning as ***they***	**express**	the speaker's scalar assessment of the

The reporting verb *address* collocates with the nouns *issues* and *questions* with a logDice value of (10.211 and 10.190, respectively), as in the following examples.

Example 6: *address + issues* in AEC

I labeled 'sociolinguistics'. These articles	**address**	***issues*** such as language and nationalism,
responses? The design of this study	**addresses**	***these issues*** in a two-pronged manner, through

Example 7: *address +questions* in AEC

allows them to do so. Specifically, we	**address**	the following questions: Which lay and

in which the researcher may	**address**	two different research ***questions*** , conducting

In the AAC, fewer reporting verbs are used with more than 90% faithfulness to the transitive clause structure. The verb */ðakara/(mention)* is the most frequently used verb in this clause structure with 99% faithfulness. Out of its total occurrences, this verb is mostly complemented by an NP (more than 70%) and less frequently by a finite-clause (less than 30%). The verb */qala/(say)* which is the second most frequently used verb in the transitive clause structure is used in this clause structure in 96% of its total occurrences. */qala/(say)* is mostly followed by a finite clause (more than 70%) and less frequently by an NP (less than 20%). The verbs */tanawala/ (deal with)* and */naqala/ (report)* are less frequent in the AAC but they are used in the transitive clause structure in more than 90% of their total occurrences. They are mostly followed by a nominal complement. The verb */tanawala/ (deal with),* Example 8, is used with an animate agent, usually a proper noun, to mean *discuss.* It is also used with an inanimate subject, e.g. the research.

Example 8: */tanawala/(deal with)* in AAC

a	الدكتور / امين سالم هذه الاسس و رد علي	**تناول**	ب جانب اضافات فرعية ([16]). و قد
b	البحث الافعال الناقصة (كان) ام الباب	**تناول**	

a) tanaːwala ʔa-ˈduktuːr Ameen Salim haðih-i

PAST-discuss-3SG DEF-doctor-NOM Ameen Salim DET-this-1F

ʔal-ʔusus

DEF-principle-PL

Doctor Ameen Salim discussed these principles

b) tanaːwala ʔal-baħθu ʔal-ʔafʕalu ʔa-ˈnaqisˁah

PAST-discuss-3SG DEF-research-NOM DEF-verb-ACC-PL DEF-copula-ACC.F

The research discussed the copular verbs

The verb */naqala/ (report),* on the other hand, is mostly associated with an animate subject, usually expressed by a proper noun, who reported an opinion or a discussion, Example 9.

Example 9: */naqala/(report)* in AAC

a	الشاطبي الخلاف في هذا التقسيم الثنائي ل	**نقل**	اي : هذا اللفظ كنز من كنوز الجنة . و قد
b	القرطبي رايا مختلفا عن ابن العربي يبين	**نقل**	و اراد ب اليدين اليمينان "([59]) . و

a) naqala ʔal-ʃatˤibij-u ʔal-χilaːf-a fi haða

PAST-report-3SG DEF-Alshatibi-NOM DEF-controversy-ACC in DET-this-1M

ʔa-ˈtaqsiːm-i

DEF-classification-OBL

Alshatibi reported the controversy over this classification.

b) naqala ʔal-qurtˤubi-u raʔi-an muχtalif-an

PAST-report-3SG DEF-Alqurtubi-NOM opinion-ACC different-ACC

Alqurtubi reported a different opinion

To sum up, it is observed that very different verbs are used in the reporting verbs semantic class in both languages with fewer examples in the AAC. Other than the verb */tanawala/ (deal with)* which is sometimes used to mean *discuss*, most reporting verbs identified as highly faithful to the transitive clause structure in the AAC are 'research type' reporting verbs (Hyland, 2000:27-29; Manan & Noor, 2014); that is they are used to merely report what is said in the literature without synthesis or criticism. This may suggest that Arab advanced learners may use fewer reporting verbs when writing academically in English. They may underuse or not use at all verbs of 'cognition' or 'discourse' type, which are commonly used to present the writers' critical thinking and evaluation, such as *claim*, and *suggest* etc.

The logical semantic relationship verbs used in the AEC with more than 90% faithfulness are *involve, contain, include, support,* and *represent.* These verbs serve an important rhetorical function in explaining the relationship between constructs of knowledge or parts of processes. For example, the verb *involve,* according to FrameNet, activates the semantic frame of 'Participation' which involves an event and participants. In many instances the event is mentioned in the subject position, as in example 10. It collocates with the noun *process* with a logDice of 8.769, as in the first and second sentences of Example 10, where it is used to explain the process that an event/ concept, such as *decolonizing* involves.

In 147 instances in the AEC, the participants are mentioned first followed by the verb in the passive voice then the event is mentioned in a prepositional phrasal that starts with *in*, as in example 11. The use of the verb *involve* serves an important function of expressing the relationship between an activity or an event and the participants.

Example 10: *involve* in AEC

schooling. This decolonizing process	**involves**	"critiquing western worldviews and
experiences. This process of construction	**involves**	the creation of a coherent, though subjective
which suggests that understanding language	**involves**	the construction of multimodal mental
Since reading comprehension necessarily	**involves**	interaction between the reader and the text in

Example 11: *involved +in* in AEC

dominant usage, which means the synonyms	**involved**	in these items are much closer in meaning and
processing in a first language.The mechanisms	**involved**	in MWE production in a second language are
sport in general. A number of parties were	**involved**	in the debate, including Marisa Da Silva (the

The verb *represent* is another logical-semantic relationship verb. It collocates with the noun phrases *progress*, *topics* and *information*, with a LogDice value of more than 7, in the object position, as in Example 12.

Example 12: *represent* in AEC

-whose discourse function was coded, 48%	**represent**	shifting topics, 51% represent continuing
agree that the development of human rights	**represents**	important progress in human history.
ospeech into information units, each of which	**represents**	a 'quantum of information ' (Halliday and

In many instances the verb *represent* collocates with the word *Figure,* with a capital initial letter, with a LogDice value of (8.096), as in Example 13. It is noted that in all instances the passive voice is used.

Example 13: *represent* + *Figure* in AEC

digging and the sand is caving in.); this is	**represented**	schematically in ***Figure*** 10 (cf. also von
, respectively. The core of our analysis is	**represented**	in Figure 27. Free permutation, which
How the alignment process operates is	**represented**	in Figure 1, which is adapted from Pickering and

Another example is the association between the verb *support* and the nouns *findings, hypothesis* and *claim* found in the sample of academic writing of this study. The verb *support* is used to highlight the relationship between *findings*, in the subject position, and *hypothesis* and *claim,* in the object position, as in Example14.

Example 14: *support* in AEC

than in the interviews. These ***findings*** do not	**support**	the hypothesis in some studies that he use is a
of both verb classes. These ***findings*** do not	**support**	the claim that the loss of the dative case leads

In the AAC, a number of logical-semantic relationship verbs are used with a high measure of faithfulness in the transitive clause structure. For example, the verb */yatatˤalabu/(demand)* is used with a nominal complement to express what a certain process, task, issue requires, as in Example 15.

Example 15: */yatatˤalabu/(demand)* in AAC

a	جهدا متزايدا ل ان الطفل : " اذا بلغ العاشرة	**يتطلب**	الاخرى "([46]). اما بعد العاشرة ف الامر
b	مهارات ل غوية جديدة : قراءة و كتابة , و	**يتطلب**	ما سبق ف ان تواصل عصر المعلومات عن بعد

a) ʔal-ʔamru jatatˤalabu dʒuhd-an mutazajdan

DEF-matter-NOM PRES-require-3SG effort-ACC increasing-ACC

The matter requires an increasing effort

b) tawasˤal-a ʕasˤr-i ʔal-maʕlumat-i ʕan buʕd

communication-ACC age-OBL DEF-information-OBL from distance

jatatˤalabu maha:ra:t-in laɣawiat-in dʒadi:dat-in

PRES-require-3SG skill-PL.ACC linguistic-ACC new-ACC

Communication in the age of distance information requires new linguistic skills.

The verb */yatadˤamnu/(include)* is used to express what a whole, such as *the book, the title* includes, Example 16.

Example 16: */yatadˤamnu/ (include)* in AAC

a	الكتاب المعلومات اللغوية المناسبة ل تلميذ	**يتضمن**	مهارات التلميذ اللغوية نموا متوازيا . - ان
b	عبارات او اشارات رمزية . ? - وضعت جامعة	**يتضمن**	توثيقية و لغوية ... و خصوصا اذا كان العنوان

a) jatadˤaman-u ʔal-kitab-u ʔal-maʕlu:mat-I ʔal-laɣawi:ah

PRES-include-3SG DEF-book-NOM DEF-information-ACC DEF-linguistic-OBL ʔal-munaːsibah
DEF-suitable-OBL

The book includes the suitable linguistic information

b) ʔal-ʕunwaːn-u jatadʕaman-u ʕibarat-in ʔaw ʔiʃarat-in
DEF-title-NOM PRES-include-3SG phrases-ACC or signals-ACC ramziːjah
symbolic-ACC

The title includes phrases or symbolic signals

Logical-semantic relationship verbs are similarly used in the AEC and the AAC. Almost the same number of verbs from this semantic class are highly faithful to the transitive clause structure (5 in the AEC and 4 in the AAC) and these verbs perform a similar semantic function of drawing the relationship between two concepts of knowledge.

5.3.3 Prepositional type 1

In the AEC, the verb *account* is used in 100% of its total occurrences in the prepositional type 1 clause structure with the preposition *for*. *Account for* is used to mean *explain.* It is used to show how the subject, such as *analysis, model, factors,* can give an explanation for the object, such as *ordering, data, the properties.* It collocates with the word *ordering* in the object position with a logDice of 10.00 and with the word *data* with a logDice of 7.94, as in Examples 17. In the subject position the prepositional verb *account for* is highly associated with the word *analysis*, with a logDice of 7.2, as in the first sentences of Example 17.

Example 17: *account for* in AEC

subject on its right. Our *analysis* cannot	**account**	*for* this one example, and we will treat it as
to a set of feature values. This analysis will	**account**	*for the obligatory occurrence of*
Heine & König's model cannot successfully	**account**	*for the ordering of R and T when the alignment*
of the three factors can successfully	**account**	*for the ordering of R and T in flagging,*
movement analysis may not fully	**account**	*for the properties of the BNP construction.*
as argument. So frequency alone cannot	**account**	*for all the data . Next, our account for the*
generally, this line of analysis might also	**account**	*for the data discussed in Section 2.1.3 above,*

In a number of instances, the verb *account for* is used with the phrase *the fact that*, see Example 18, to give the meaning of recognizing that fact, including it in a further discussion and providing explanation for it.

Example 18: *account for+ fact* in AEC

This aspect of the Rwandan macro policy also	**accounts**	***for*** the ***fact*** that, as mentioned earlier,
characterise these expressions. We will also	**account**	***for*** the ***fact*** that same is not semantically

The prepositional verb *refer to* is also the most frequently used combination in this clause structure in the AEC. It is used with the preposition *to* in this clause structure in 96% of its total occurrences. It is mainly used to define a certain term as in Example 19.

Example 19: *refer to* in AEC

The ***term***	**refers**	***to*** anything that relates to our name and surname
The term	**refers**	, accordingly, ***to*** the functional value of a

The prepositional verb *contribute to* is another prepositional verb that is used in the AEC to express logical-semantic relationships with a high frequency (217) and a high measure of faithfulness (81%). The verb *contribute to* is an important lexical item that points to the novelty of the research and its contributions to the body of knowledge/academic debates. It is used with subjects, such as *article, paper, study* to demonstrate how they add to the *literature*, the *understanding*, the general *knowledge,* see Example 20. Such usage is very useful for learners because it helps them express the importance of their writing, a function that they may need in their future writings in the field of applied linguistics.

Example 20: *contribute to* in AEC

of the coming-out genre. This ***article***	**contributes**	***to*** literature on narrative and genre by
multidisciplinary interest. This ***article***	**contributes**	***to*** this line of research by considering

It can be noted that the prepositional verbs that are most faithful to the prepositional type 1 clause structure, in the AEC, are verbs that express a logical-semantic relationship. They are used to explain the relationship between two constructions of knowledge, e.g. the relation between *analysis* and *data*. In the AAC, the prepositional verb /*dalla/ (indicate)*

ʕala (on), Example 21, which is used in the prepositional type 1 clause structure with a measure of faithfulness of more than 90% expresses a logical semantic relationship. It is used in the corpus to explain the relationship between two constructions of knowledge.

Example 21 : */dalla/ (indicate) ʕala (on)* in AAC

a	على المعنى .	تدل	و هي اشارات

a) hija iʃaːraːt-un tadulu ʕalaː ʔal-maʕna

3SG.F signs-NOM PRES-refer-3SG to DEF-meaning

They are signs that refer to the meaning

The prepositional verb */ʔʃara/ (point to/ mention) ila (to)*, Example 22, which is also faithful to the prepositional type 1 clause structure, is used in the AAC as a reporting verb. It is used with an animate subject, mostly a proper noun and has the meaning of point to/refer.

Example 22: */ʔʃara/ (point to/ mention) ila (to)* in AAC

a	الباحث الى جهد المجمع المصري	يشير	
b	الدكتور علي محمد الى عدم قدرة	يشير	

a) juʃiːru ʔal-baħiθu ʔila dʒuhdi ʔal-madʒmaʕ-i

PRES-refer-3SG DEF-researcher-NOM to effort-OBL DEF-community-ACC

ʔal-masˁri

DEF-Egyptian

The researchers refer to the effort of the Egyptian community.

b) juʃiːru ʔa-ˈduktuːr ʕali Mohammed ʔila ʕadami qudrati

PRES-refer-3SG DEF-doctor-NOM Ali Mohammed to Non DEF-ability-OBL

Doctor Ali Mohammed refers to the inability…

The third of the most faithful verbs to the prepositional type 1 clause structure is */yaħtadʒu/(need) ila (to)*, Example 23. This verb has the meaning of *need* in English, but unlike the verb *need* in the AEC which is more faithful to the transitive clause structure and is followed by either a noun phrase or an infinite clause *to-infinitive*, the verb */yaħtadʒu/(need)* in the AAC collocates with the preposition *ila* followed by an NP in 94% of its total occurrences. It is used to express a logical semantic relationship between two constructions of knowledge.

Example 23: */yaħtadʒu/(need)* in AAC

a	الى علم العربية القديم	**يحتاج**	و فهم اللغة
b	الى مزيد بيان وشرح	**يحتاج**	و هذا الامر

a) fihmu ʔal-luɣat-i jaħtaju ʔila
understanding-NOM DEF-language-OBL PRES-need-3SG to
ʕilmi ʔal-ʕarabijat-i ʔal-qadiːm
science DEF-arabic-OBL DEF-old

Understanding language needs to (requires) the old language science

b) haða ʔal-ʔamr-u jaħtaju ʔila maziːd-i bajan-in
DET-this-3SGM DEF-matter-NOM PRES-need-3SG to more clarification-OBL
wa ʃarħ-in
and explanation-OBL

This matter needs (to) more clarification and explanation

In general, it can be concluded that prepositional verbs that are highly faithful to the prepositional type 1 clause structure in the sample of academic writing in the field of applied linguistics of both languages are mainly logical-semantic relationship verbs which emphasizes the importance of this semantic class in academic writing.

5.3.4 Complex copular

As far as the complex copular clause structure is concerned, none of the verbs investigated in the AEC is more than 90% faithful to that clause structure meaning that this structure is less restricted and can accommodate a wide range of verbs. In the AAC, three verbs are faithful to that clause structure. The two activity verbs */dʒaʕala/(make/ create)*, Example 24, and */samma/ (call)*, Example 25, are used with two objects that can be equated with the subject and its predicate. The mental verb */yuʕtabaru/ (consider)*, Example 26, which means *consider* is used in this clause structure in 99% of its total occurrences. The verb */yuʕadu/(consider/ is no longer/ count)* which also has the meaning of *consider* is very frequently used in this clause structure with a measure of faithfulness of 76%.

Example 24: */dʒaʕala/(make/ create)* in AAC

a	ها لغة الدراسة في جميع المواد	**جعل**	

a) dʒaʕala-ha luɣat-a ʔa-ˈdirasat-i fi dʒamiːʕi ʔal-mawad-i

PAST-make-3SG-[he]-it language-ACC DEF-study-OBL in all DEF-subject-PL

He made it the language of study in all subjects.

Example 25: */samma/ (call)* in AAC

a	طريقة القواعد و الترجمة (Grammar Translation	**تسمى**	الام او القومية . و هذه الطريقة

a) haðihi ʔa-ˈtˤariːqat-i tu-samaː tˤariːqat-a

DET-this-3SGF DEF- method-OBL PRES.PASS-called-3SG method-ACC

ʔal-qawaːʕid-i wa ʔa-ˈtardʒamat-i

DEF-grammar-OBL and DEF-translation- OBL

This method is called grammar and translation method.

Example 26: */yuʕtabaru/ (consider)* in AAC

a	فرديناند دي سوسور الاب الحقيقي ل السانيات .	**يعتبر**	

a) j-uʕtabaru Ferdinand de Saussure ʔal-ʔab-a ʔal-ħaqiqi-a

PRES.PASS-consider-3SG Ferdinand de Saussure DEF-father-ACC DEF-true-ACC

li-lisaːniːat-i

for-linguistics

Ferdinand de Saussure is considered the true father of linguistics.

5.4 Lexico-grammatical analysis of selected clause structures

This section focuses on the use of verb-noun collocations in phrasal verbs type 2, phrasal prepositional type 1,2, and 3 clause structures in academic English writing. These clause structures are selected for a detailed analysis for a number of reasons (see also 2.5.5.6). These reasons include the nonexistence of such structures in Arabic, which is the first language of the target students, which makes these structures potentially more error prone for Arabic learners. Another reason is the reported general difficulty of these clause structures for ESL students (e.g. Paquot & Granger, 2012:133).

Furthermore, the initial analysis of the distribution of verb complementation clause structures clearly shows that phrasal verbs are used frequently in applied linguistics academic writing. More than 20% of the total of types identified for the most frequent 100

verbs in AEC are phrasal verbs. Additionally, phrasal verb type 2 is the fourth most frequently used clause structure in AEC and phrasal prepositional type 1 is the sixth most frequently used clause structure, when verb types are considered. The results of this study contradict the long belief in the rarity of phrasal verbs in formal contexts (Biber et al., 2002; Biber et al., 1999; Liu, 2011) and show that the use of phrasal verbs is one of the important characteristics of current academic English writing in the field of Applied Linguistics. It may be indeed a recent tendency, which has been so far overlooked in research.

Studies on English phrasal verbs mainly focus on listing and defining them, little is said about the argument structures and the semantic roles involved, a gap which this research aims to address. Additionally, while previous research on phrasal verbs focuses mainly on general use (Garnier, 2016; Garnier & Schmitt, 2015), this section shows how phrasal verbs are used in academic writing and compares this use with previous research on phrasal verbs in general English.

This section also includes a detailed analysis of prepositional type 3 clause structure. In this clause structure, the verb and its complementation form a fixed expression, such as *take care of,* and *play a role in.* Although this clause structure is found in Arabic, it is not that common. Also due to its rather idiomatic collocational nature, it may be difficult for learners. Therefore, it is included in this section for its instances in the AEC to be closely analysed. The above general results show the high frequency of this clause structure in the AEC, 517 tokens and 25 types are used in this clause structure. Despite their frequent usage in the sample of academic writing investigated here, to the best of the research's knowledge, this clause structure has not been investigated in detail in previous research on academic writing.

The following analysis includes verb types with a raw frequency of more than 5. This is because 5 occurrences are not sufficient to detect a pattern of usage. For convenience of analysis and to provide insights into fine grained semantic differences between phrasal verbs, phrasal verbs which consist of the same lexical verb combined with different particles are presented together. For example, *take up* (56), *take on* (32), and *take over* (9), are discussed together.

5.4.1 Phrasal type 2

Phrasal type 2 has the following syntactic clause structure SVOd (NP VP AdvPart NP), in which the verb phrase is preceded by a noun phrase and complemented with another

noun phrase. The initial analysis identified 41 verb types used in this clause structure (Table 5-4). This detailed analysis focuses on 8 of them: *take up* (56), *take on* (32), *take over* (9), *set up* (31), *set out* (13), *make up* (21), *find out* (17), and *work out* (12).

Table 5-4: Phrasal type 2 in AEC

	Phrasal verb	**Raw**	**Nom**	**Faith%**		**Phrasal verb**	**Raw**	**Nom**	**Faith%**
1	*take up*	56	6	6	22	*take away*	2	0	0
2	*take on*	32	3	3	23	*take forward*	2	0	0
3	*set up*	31	3	12	24	*take in*	2	0	0
4	*make up*	21	2	2	25	*turn back*	2	0	1
5	*find out*	17	2	2	26	*call out*	1	0	0
6	*set out*	13	1	5	27	*get cross*	1	0	0
7	*work out*	12	1	4	28	*get through*	1	0	1
8	*take over*	9	1	1	29	*give away*	1	0	0
9	*start off*	5	1	2	30	*give in*	1	0	0
10	*follow up*	4	0	0	31	*give up*	1	0	0
11	*start out*	4	0	2	32	*lead on*	1	0	0
12	*get in*	3	0	1	33	*mark out*	1	0	0
13	*get out*	3	0	1	34	*set forth*	1	0	0
14	*turn off*	3	0	1	35	*take aback*	1	0	0
15	*turn up*	3	0	1	36	*think over*	1	0	0
16	*draw up*	2	0	0	37	*turn down*	1	0	0
17	*move away*	2	0	1	38	*turn on*	1	0	0
18	*set apart*	2	0	1	39	*turn round*	1	0	0
19	*set aside*	2	0	1	40	*write down*	1	0	0
20	*set off*	2	0	1	41	*write out*	1	0	0
21	*show off*	2	0	0					

TAKE UP (56)

The most frequent phrasal verb of the top 100 verbs in the AEC is *take up*. In general use, in the BNC and COCA, according to Garnier and Schmitt (2015), the most frequent senses of *take up* are, in order of frequency:

1. Use a particular amount of space, time or effort

2. Discuss or deal with (issue, idea, matter)
3. Starting to do a particular job or activity.
4. *Take up* also has a literal meaning of 'grasping an object, often moving it from a lower to a higher position'.

The first three metaphoric senses of *take up* are found in academic writing but in a different order. A close semantic analysis of the concordance lines of *take up* in the sample of English academic writing provides insights into the types of noun phrases associated with this verb and its patterns of meaning. The subject of this phrasal verb may be animate, inanimate or unknown, because the passive voice is used. Animate subjects include *writers, interviewees, subjects;* pronouns, such as *they, I* and *we;* and proper nouns, such as *Kibort.* Inanimate subjects include nouns such as *paper, article, chapter, section.* The object position is mostly filled up with words such as *questions, negotiations, argument,* and *challenge.* This pattern of use is detected in 23 instances out of the total 56, therefore, it can be concluded that in English academic writing the second sense of the list above is more common and that the nouns that fill in the object position to be academically *taken up* are nouns of argument or dispute. The subject is an arguer that *take up an argument, a question*, or *a negotiation* as in the following examples:

Example 27: *take up* (discuss/ to deal) in AEC

and Big Pit coal mine in South Wales. The paper	**takes up**	again the negotiation of place and experience
local school dynamics. The following section	**takes up**	this argument in more detail by analyzing
context of multimodal practices? This article	**takes up**	these questions by examining the perspectives
not be limited to writing or speaking). I will	**take up**	each of the strategies in turn in order to

In 20 instances, the object position of *take up* is also filled up with words like *the role, the position, the idea*. In this case, the subject is either animate or inanimate agent that adopt a position.

Example 28: *take up* (a position) in AEC

' (Calver 1946: 323). A similar idea is	**taken up**	by Dowty (1975), who assumes that the contrast
. Whether readers subscribe to the position	**taken up**	in Mehdi Riazi's paper or not, it raises issues
absolute constructions confirms the stance	**taken up**	by reference grammars for earlier periods in
practice (Hanks 1990, 2005) is used to	**take up**	a position in the social field of stand-up

The subject may also take the role of an activity that takes a length of time or space, in sentences like:

Example 29: *take up* (activity) in AEC

a considerable amount of the time is	**taken up**	in translation back and forth. From the point of
gained in importance quite spectacularly,	**taking up**	22% of all instances of the absolute

The first two semantic frames of (arguer- take up- an argument) and (agent-take up- a role) are more common in academic writing (43 instances out of 56) in the AEC than the third associated with time and space, which in turn is the most frequent sense in general use (Garnier and Schmitt, 2015).

TAKE ON (32)

The second most frequently used phrasal verb is *take on.* According to Garnier and Schmitt (2015), in general use, *take on* has the meaning of:

1. Undertaking or handling a role, a task, a responsibility, a problem, or an issue
2. Acquire or assume as a quality of one's own, a meaning, a colour, or a shape.

In the sample of academic writing studied here, the subject position is filled up by both animate and inanimate nouns. Examples of animate subjects include proper nouns, such as *Frida &Dandi* and *Rona,* pronouns such as *she, I* and *we*. Nouns such as *applicants, analysts* are also used as animate subjects. Examples of inanimate subjects include *forms, words,* and *variables.* The word *meaning(s)* occurs in 8 concordances in the object position. Words of importance like *salience, significance, roles, functions* are examples of other words used in the object position. It can be noted that in many examples the subject has the role of a recipient of an attribute or a theme. For example, in the following sentence, the subjects *words, forms, multimodality,* and *crying* receive *meaning*s (an attribute) or *topics* and *forms* (theme).

Example 30: *take on* + inanimate subject in AEC

are recurring strings of words which have	**taken on**	more specific meanings than the composition
have demonstrated how linguistic forms can	**take on**	complicated social meanings that span
concepts is that multimodality tends to	**take on**	a huge range of topics, encompassing
domestic helpers. Crying in FDH narratives	**takes on**	many different forms and in order to

Even when the subject is animate, it still has the role of a recipient. In the following sentences, the subjects *I, the applicants, health care providers* receive or accepts *a role* or *an adventure.*

Example 31: *take on* + animate subject in AEC

et al. 1992). In the sharing sessions, I	**take on**	the role of the interviewer who asks
make the applicant feel comfortable about	**taking on**	an exotic adventure in an unfamiliar
Health care providers negotiate their professional identities and	**take on**	particular roles in response to the

Generally, it may be noted that in all concordance lines of *take on*, the subject is mentioned, as all examples are in the active voice. It also can be noted that when the agent is animate it is more likely to *take on* a role/ a stance/ an adventure. However, inanimate subjects collocate with *meanings* and *forms*. Both meanings are similar to the two meanings found in general use.

TAKE OVER (9)

In academic writing in the AEC, the phrasal verb *take over* is not as frequent as the verbs *take up* and *take on.* However, in the PHaVE list (Garnier and Schmitt, 2015), it is reported to be more frequent than *take up* with the meaning of:

1. Gain control, management, or possession of something or somebody (a task, job, political party, or organisation)

The subject of *take over* in academic writing is inanimate, e.g. *morph, construction, preposition* and animate, e.g. *I*. The object is a position or role. According to FrameNet, the verb *take over* activates the semantic frame of 'Change_of_Leadership', this frame involves a new leader, an old leader and a function. This semantic frame is activated by the verb *take over* both in general and academic use. In Example 32, the new leaders are *a single morph* and *prepositions* which *take over* the *function/ the role* of the old leader *a combination of two morphs, morphological case.*

Example 32: *take over* in AEC

portmanteau (Stump 1993): a single morph that	**takes over**	the role played by a combination of two morphs in
informed by the assumption that prepositions	**took over**	the function of morphological case in Middle

A subject might *take up a role, take on a role* or *take over a role.* With *take up*, the subject is an agent who starts doing that role, with *take on,* the subject is a recipient who accepts the role, and with *take over,* the subject is the new leader who replaces an old leader. *Take up* is frequently used in stative passive where the subject is unknown, with few examples of regular passives with the subject mentioned in the *by*-phrase*;* whereas, *take on* and *take over* are mostly used in the active voice. This indicates that with the verb *take up,* the activity/ the topic is more important than the agent but with *take on* and *take over,* the subject which plays the role of the recipient or the new leader is more important.

SET UP (31)

In general use, *set up* is ranked as the eleventh of the most frequent phrasal verbs listed in the PHaVE list (Garnier and Schmitt, 2015). It has two senses:

1. A metaphoric more common sense which is to establish or create something arrange for something to happen or exist, and
2. A literal less common sense which is to place something in a particular spot or position.

This phrasal verb is the third most frequent verb in the sample of this study. In many instances of this phrasal verb, the subject is animate. Pronouns, such as *he, she,* as well as other nouns which refer to animate entities such as *the therapist, the nurse, people, the IR* are used in the subject position. In some examples the subject is unknown in a passive structure. Few examples are found with an inanimate subject such as *the event* and *the group.* In the object position, words like *group, a typology, the need, the contrast, the dichotomy, the relation* are used. These are abstract entities related to the academic process of grouping and finding patterns, or common or distinctive features. With these objects the phrasal verb *set up* has the metaphoric sense of create, similar to its more common sense in general use. With the sense of create, the verb *set up* activates the semantic frame of 'Creating' which involves an agent/ creator and a theme/ a configuration/ a created entity. In Example 33, the subjects *we* and *event* are agents that create the themes *typology,* and *relation.*

Example 33: *set up* (create) in AEC

of morphotactic systems, we will start by	**setting up**	a canonical typology of morphotactics.
corpus a cause–effect relation (Impact) is	**set up**	by event (e.g. hurricane) + (the) subsequent

Another smaller set of concrete objects includes *the screen, the video camera, the booth;* in these examples, *set up* has the sense of literally placing an item in a certain location. This sense activates the semantic frame of placing which involves an agent and a theme, as well as a goal which refers to the place where the theme is placed. In Example 34, unknown agents place the themes *booth* and *screen* at *speaker's corner* and *at the front of the room.*

Example 34: *set up (place)* in AEC

at a "Speaker's Corner" booth that was	**set up**	to enable them to share their experiences.
table. Datashow equipment and a screen are	**set up**	at the front of the room. Philip has bee

In the following examples, the agents are *she*, and *the IR*, the themes are the reflexive pronouns *herself, her/himself.* The goals in these examples are not a place or a location but a position or a role that the agent place her/himself in.

Example 35: *set up (take a role)* in AEC

in a similar situation). In doing so, she	**sets**	herself up as a co-decision maker. However
political activities. The IR, meanwhile,	**sets**	her/himself up as a representative/intermediary

The use of *set up* in academic writing resembles its use in general language. *Set up* is more frequently used in its metaphoric sense in relation to *setting up* arrangements, classifications and typology.

SET OUT (13)

According to the PHaVE list (Garnier and Schmitt, 2015), *set out* has three meanings, of which the first is the most commonly used in general language:

1. To start doing or working on something.
2. Start a journey
3. Explain or present something clearly, especially officially and in writing

However, when *set out* is used in academic writing, it more frequently has the third sense which is to explain or present something clearly. Set *out* is used in the passive voice in most of the examples of its transitive use investigated here, and the subject is unknown, only in three examples the subject is mentioned in a *by*-phrase. The pronoun *we* is used as

a subject in two examples. *A model, a plan, the criteria, an agenda,* and *the results of an evaluation* are examples of the nouns that fill in the object position. It is noted that in most instances a location for setting out the object is mentioned, such as *in this section, in line 5, in figure.* Therefore, it can be concluded that *set out* has more of a metadiscursive function (Hyland & Tse, 2004) referring to elements in the text.

Example 36: *set out (explain)* in AEC

, and the results of this evaluation are	**set out**	in a document, the Language Analysis
applied in the context of language analysis is	**set out**	in Figure 1. Issues arise in the process

While *set out* is mostly used in general language to mean start an activity, its most common meaning in academic writing is to explain clearly. The general use of *set out* is found in academic writing but it is less frequent, as in:

Example 37: *set out (start an activity)* in AEC

this line of thinking, the present study	**sets out**	to examine alignment in L2 learning with the
little described languages. In this paper we	**set out**	to provide some answers and advice about these

A subject may *set up a classification* or *set out a model/ the criteria*, as in the following examples. With *set up*, the subject is the creator of an entity (*the classification*). However, with *set out*, the subject may or may not be the creator of the entity *the model/ the criteria*, but, most importantly, the subject is the agent that provides a detailed explanation of that entity in a certain place in the text.

Example 38: *set up* + *set out* in AEC

Based on corpus data, we have	**set up**	a broad classification that captures the range
of English NP structure In this section we	**set out**	our model of the basic functions coded by the

MAKE UP (21)

The phrasal verb *make up* is the fourth most frequent phrasal verb in the data at hand. In general language use reported in the PHaVE list (Garnier and Schmitt, 2015), it has three main senses:

1. To form the whole or an amount of an entity and
2. Compensate for sth lacking, lost or missed
3. To make a decision

The analysis of the use of this phrasal verb in AEC shows that the first sense is more common in academic writing. Only two examples, presented below, are detected in which *make up* is associated with making a decision in the phrase *make up someone's mind,* and one of them is within quotation marks which is a citation reflecting general spoken language use.

Example 39: *make up your mind* in AEC

have a comparison of candidates' rivals to	**make up**	their minds to vote. The data for analysis are
"listen for yourself with fresh ears and	**make up**	Your own mind". To help the audience determine

In sentences where *make up* has the literal sense of forming the whole or an amount of an entity, the subject is inanimate, such as *it, the elements, categories, constructions.* The object position is occupied with nouns of amount such as *the bulk, 2%, proportion, the bigger unit,* as in the following examples:

Example 40: *make up (form an amount of sth)* in AEC

a situation noun. Because this group	**makes up**	a substantial portion of the data, examples of
4 per cent of the total, while the SVX pattern	**makes up**	more than four-fifths of all clauses. At this

It can be generally said that the use of *make up* in academic writing is similar to its use in general English. Its literal sense is more frequent in written and spoken language use.

FIND OUT (17)

In general use, the phrasal verb *find out,* means to discover or obtain knowledge about something. It activates the semantic frame of 'Becoming_aware', which involves a cognizer, means and phenomenon/topic as core elements. Topic is the general field within which the phenomenon fell. This meaning and the semantic frame are also found in academic writing. The cognizer role is expressed mostly by pronouns, such as *I, we,* and *he,* a noun phrase such as *the therapist, people, readers.* The role of topic/ phenomenon is mostly expressed as a question word, such as *whether* and *what.* In some examples, such

as the ones presented below, words like *several sequences*, *external website*, *a question* are used to express the means which the cognizer uses to find out about the topic.

Example 41 : *find out* in AEC

several film sequences from this project, we	**find out**	that this textual composition is a specific
.co.uk's external website in order to '	**find out**	more' and know 'the causes and treatment'.
starts with a question of the therapist to	**find out**	what the client had done to cope with the stress

WORK OUT (12)

In general use, *work out* has three meanings:

1. To plan, devise or think about something carefully or in detail
2. Exercise in order to improve health or strength
3. Happen or develop in a particular way

However, the analysis of this verb in AEC reveals that in most of its occurrences the phrasal verb *work out* means to understand. In this sense, it is similar to the phrasal verb *find out.* Therefore, it activates the same semantic frame of becoming aware which involves a cognizer, and a phenomenon/topic as core elements. Examples of phenomenon/topic include *the meaning's,* question words such as *what* and *how*, as in the following sentences:

Example 42: *work out (understand)* in AEC

in the person's pragmatic understanding	**working out**	what people mean by non-literal expressions,
to most linguistic fieldworkers: How can I	**work out**	the meanings of unfamiliar words and

In some examples, the verb *work out* is used in its more general sense to mean plan, as in Example 43. With this sense the semantic roles are different. Rather than a cognizer and a phenomenon, the verb is associated with the roles of an agent and theme. In the following examples, the agent needs to plan/devise *a typology* or *a classification.*

Example 43 : *work out (plan)* in AEC

'Towards a typology of grammatical features',	**works out**	a typology informed by the descriptions of a
discussion, a simple classification can be	**worked out**	for English premodifying present participles

In general use, the verb *work out* most frequently means to plan or devise something. This sense is less frequent in academic writing with the meaning of understand being more common.

5.4.2 Phrasal prepositional type 1

Phrasal prepositional type 1 has the following syntactic clause structure SVOp (NP VP AdvPart PP), in which the verb phrase is preceded by a noun phrase and complemented with a prepositional phrase. The initial analysis identified 25 verb types used in this clause structure (Table 5-5). This detailed analysis focuses on 2 of them: *come up with* (10) and *move away from* (7).

Table 5-5: Phrasal prepositional type 1 in AEC

	Phrasal prepositional type1	**Raw**	**Norm**	**Faith%**		**Phrasal prepositional type1**	**Raw**	**Norm**	**Faith%**
1	*come up with*	10	1	1	14	*start off with*	3	0	1
2	*move away from*	7	1	1	15	*start out with*	3	0	1
3	*come back to*	5	1	0	16	*write back to*	3	0	1
4	*go along with*	5	1	1	17	*come out with*	2	0	1
5	*follow up on*	4	0	1	18	*come up against*	2	0	0
6	*look forward to*	4	0	0	19	*get down with*	2	0	0
7	*refer back to*	4	0	1	20	*get away with*	1	0	1
8	*develop out of*	3	0	1	21	*get out of*	1	0	1
9	*get on with*	3	0	1	22	*get up to*	1	0	0
10	*go out of*	3	0	1	23	*go up to*	1	0	0
11	*lead up to*	3	0	1	24	*move along with*	1	0	0
12	*look back at*	3	0	1	25	*begin out of*	1	0	3
13	*look up at*	3	0	1					

COME UP WITH (10)

In general use, *come up with*, is used to mean bring forth or produce. It activates the semantic frame of ‘Coming_up_with’ which involves a cognizer who comes up or conceptualizes an invention or an idea.

In the AEC, *come up with,* also activates the same semantic frame. In all examples, the verb is used in the active voice mostly preceded with an animate cognizer, such as *we, they, you, linguist, the speakers*, except for one example where the subject is an inanimate *frequency list.* The inventions that the cognizer *comes up with* include products of thinking and academic engagement such as *questions, interpretations, explanation, terms, understanding*, as in:

Example 44: *come up with* in AEC

problem of a similar kind – linguists tend to	**come up with**	different terms for what is essentially
line with the above contentions, we have	**come up with**	the following research questions
rhetorical functions Matthiessen & Thompson	**come up with**	a new, more specialized understanding

MOVE AWAY FROM (7)

Move away from generally refers to one's physical movement from one place to another, and more commonly used when one stops living in a place and goes to live in another place. However, in the AEC, *move away from* is used metaphorically with the meaning of shifting one's ideas or beliefs. The subject moves away from *the norms, thinking, the view, the perspective.* The subject may be animate, such as *we, teachers,* or inanimate as in Example 45. Therefore, in academic writing the metaphoric use of this verb dominates.

Example 45: *move away from* in AEC

collective illusions (Carfantan, 2003), we are	**moving**	away from cognitive reality to social reality
dominant racial ideology. So she urges us 'to	**move**	away from thinking of racism as entirely
sensitive pedagogies" that can help TESOL	**move**	away from the Western norm embedded in
suggest that the use of participles has been	**moving**	away from the core functions of nominal

5.4.3 Phrasal prepositional type 2

Phrasal prepositional type 2 has the syntactic clause structure of SVOdOp (NP VP NP AdvPart PP), in which the verb phrase is preceded by a noun phrase and complemented with a noun phrase, an adverbial particle and a prepositional phrase. The NP after the verb

is often the pronoun *it* which forms part of a fixed expression. The initial analysis identified 6 verb types used in this clause structure (Table 5-6). Some of these verbs are movement verbs but are used metaphorically to indicate back and forth movements of ideas and beliefs (*turn away*, *follow*, *lead away*). The use of these verb combinations gives the image of academic writing as a space in which ideas, concepts and beliefs are being moved, shifted etc. This section covers the verb *made up of* because its frequency is more than 5.

Table 5-6: Phrasal prepositional type 2 in AEC

	Phrasal prepositional type 2	**Raw**	**Norm**	**Faith%**
1	*made up of*	16	2	2
2	*turn it away from*	2	0	1
3	*follow it up with*	2	0	0
4	*lead the discourse away from*	1	0	0
5	*set it apart from the views*	1	0	0
6	*take it out of its context*	1	0	0

MADE UP OF (16)

In all instances, *made up of* is used in the passive voice. Thus the subject is unknown. In general use, *make up of* means to form the whole or an amount of an entity. It activates the semantic frame of Creating which involves a creator, and a created entity as core elements. The Creating semantic frame also involves the mentioning of the components of which something is created as a non-core element. In this phrasal-prepositional verb *made up of*, the creator is not mentioned because the passive voice is used, however, the created entity is mentioned as well as the components of which it is made, as in the following examples:

Example 46: *made up of* in AEC

The extracts were all	**made**	up of complete sentences. The fifty-four
The BNC is	**made**	up of 10 per cent spoken and 90 per cent

The phrasal prepositional verb *make up of* has the same meaning, in general and academic use. However, in academic writing it is more frequently used in the passive voice.

5.4.4 Phrasal prepositional type 3

The syntactic structure of phrasal prepositional type 3 is similar to that of phrasal prepositional type 2; SVOiOp (NP VP NP AdvPart PP). The only difference is that the NP after the verb is an indirect object affected by the verb. The initial analysis identified 5 verb types used in this clause structure with a frequency less than 5 (Table 5-7). Therefore, they are not analysed here. Like phrasal prepositional type 2, some verb combinations are used to express metaphoric movements within the 'space' of academic writing.

Table 5-7: Phrasal prepositional type 3 in AEC

	Phrasal prepositional type 3	**Raw**	**Norm**	**Faith%**
1	*set them/sth apart form*	4	0	2
2	*get sth/someone back on*	2	0	1
3	*take sth out of*	2	0	0
4	*take us away from*	1	0	0
5	*give herself up to*	1	0	0

5.4.5 Prepositional type 3

The syntactic structure of prepositional type 3 is SVOdOp (NP VP NP PP/ NP VP PP NP) in which the verb phrase is preceded by a noun phrase and complemented with a noun phrase with which it forms a fixed collocation, such as *take care of*, then a prepositional phrase. Prepositional type 3 also includes another structure in which the verb is preceded by a noun phrase and complemented with a prepositional phrase with which it forms a fixed expression, such as *take into account,* then another noun phrase. The initial analysis identified 29 verb types used in this clause structure (Table 5-8). The analysis in this section covers 17 of these verb types that have more than 5 occurrences.

Table 5-8: Prepositional type 3 in AEC

	Prepositional type 3	**Raw**	**Norm**	**Faith %**		**Prepositional type 3**	**Raw**	**Norm**	**Faith %**
1	*play a role in*	134	14	47	16	*turn attention to*	8	1	3
2	*give rise to*	80	8	8	17	*take responsibility for*	7	1	1
3	*take sth into account*	62	6	6	18	*come to grips with*	5	1	1

	Prepositional type 3	Raw	Norm	Faith %		Prepositional type 3	Raw	Norm	Faith %
4	*make use of*	44	4	4	19	*take care of*	4	0	0
5	*draw attention to*	43	4	10	20	*take stock of*	4	0	0
6	*make reference to*	13	1	1	21	*get rid of*	3	0	1
7	*take part in*	13	1	1	22	*make fun of*	3	0	0
8	*focus attention on*	12	1	2	23	*call to mind*	3	0	1
9	*make sense of*	11	1	1	24	*call attention to*	2	0	1
10	*call sth into question*	11	1	4	25	*take control of*	2	0	0
11	*give way to*	10	1	1	26	*take hold of*	2	0	0
12	*take advantage of*	10	1	1	27	*take precedence over*	2	0	0
13	*take sth into consideration*	10	1	1	28	*come to terms with*	1	0	0
14	*play a part in*	9	1	3	29	*get hold of*	1	0	0
15	*take account of*	8	1	1					

PLAY A ROLE IN (134)

This is one of the most frequent combinations in the AEC. Like in general use, *play a role in* is used in AEC; to indicate that something is involved or has an effect on something else. However, it adds a sense of significance to that effect. The subject of such combination may be an animate, such as *they*, *friends*, *interlocutors*, but more frequently the subject is inanimate, such as *language learning, students' L1, media, discourse, affordances.* The object of the preposition *in* is usually a process expressed either by *-ing* form, such as *processing, learning, determining, resolving, limiting, realising, shaping;* as in example 47, or in the noun form ending with *-tion*, such as *formation, determination, acquisition, grammaticalization,* as in example 48.

Example 47: *play a role in* +process in AEC

Primus' assumption that thematic relations	**play**	a more important role in determining the
textual – that is, the view that the speaker	**plays**	an active role in organising and structuring

Example 48: *play a role in* + noun-tion in AEC

2003), I do not think that weight should	**play**	any role in the determination of the order
influential role that recruitment websites	**play**	in the discursive construction of our field

The role played is mostly modified by adjectives highlighting significance such as *important, significant, increasing, crucial, pivotal, privileged, n `o small, key,* as in the following examples:

Example 49: *play a role in* + adjective in AEC

Over the last two decades, agreement has	**played**	a central role in shaping a variety of
Gordon 2009), I have shown that epistemics	**plays**	an understudied but important role in these

PLAY A PART IN (9)

play a part in is used similarly *to play a role in*. It is complemented with a process and the part played was described as *crucial, important, significant,* as in the following examples:

Example 50: *play a part in* in AEC

later than the CV portion and therefore	**plays**	a less crucial part in language comprehension
increased visual information we receive will	**play**	a significant part in shaping linguistically

It is noted that while in the BNC, the verb *play* collocates with *role* and *part* almost similarly with a frequency of 3,516 for *role* and 3,176 for *part* and LogDice value of *role* (10.855) and *part* (10.091). However, in the AEC, *play a role* is more frequently used than *play a part,* as the first has the total occurrences of 143 while the latter has the total occurrence of 9 only. Although both can be used synonymously in general use, in academic writing in applied linguistics *play a role in* seems preferred probably because of the sense of significance that *play a role* adds to the effect described, whereas the word *part* in *play a part in* indicates that the effect may be partial. Therefore, if a researcher wants to say that something is important and has a significant, not a partial effect, *play a role* offers a better choice.

GIVE RISE TO (80)

In general language use, *give rise to* means to cause something to happen. In all instances of *give rise to* in the AEC, the subject was inanimate, such as *knowledge, construction, factor, compounding, the values, the effects, grammar.* This inanimate subject, like in general use, causes or allows another theme to appear or happen. The theme can be an increased result modified by *more*, the same or a new result, as in the following examples:

Example 51: *give rise to* in AEC

the L1-based contextual knowledge would	**give**	rise to more interlingual errors. These
constraints on L2 use. Similar situation models	**give**	rise to similar internal contexts, which
regularities can "[trigger] changes that	**give**	rise to new regularities" (De Smet 2012b

The use of this combination *give rise to* plays the functions of explaining the reasons or causes of a certain theme, it is a good alternative for the single-word verb *cause.*

GIVE WAY TO (10)

In general use, *give way to* means to allow something to replace something else, especially if it is cheaper, or better. However, in AEC, *give way to* is used in the same way as *give rise to* and means mostly causing something to happen. It is used with inanimate subjects, such as *insignificance, effect, time, the voice of the journalist,* which causes *classificational doubt, extravagances, other meanings* and *the voice* to occur, as in the following examples:

Example 52: *give way to* in AEC

in (4). In some instances, insignificance	**gives**	way to classificational doubt; in (5),
effect that the originals no longer produce,	**gives**	way to some ludicrous extravagances: It
and contexts (including cases where time	**gives**	way to other meanings). In other words,
one another: the voice of the journalist	**gives**	way to the voice of the "special agent"

In the BNC, the verb *give* collocates with *rise* with a LogDice of (8.887) and with *way* with a LogDice of (8.138). So both seem to have high collocability with the verb *give.*

However, in the AEC, *give rise to* is more frequently used than *give way to* which may indicate that its use is a characteristic of academic writing.

TAKE STH INTO ACCOUNT (62)
TAKE STH INTO CONSIDERATION (10)
TAKE ACCOUNT OF (8)

These three multi-word verbs have a similar meaning, they mean to consider or remember something when deciding on or designing something. They are listed as synonyms in phrasal dictionaries. However, the analysis of the concordance lines in which they are used in the AEC, reveal some fine differences between these multi-word verbs. *Take sth into account* is the most frequently used verb combination, it seems to be preferred in academic writing, as it is more flexibly used with both animate subjects, such as *we, Sebba,* and inanimate subjects, such as *studies, approach.* It is also used both in active and passive voice, as illustrated in the following examples.

Example 53: *take into account* in AEC

between 48.8% and 51%, depending on whether we	**take**	into account the answers of all the listeners
impossible in MHR texts, Sebba (2002) does not	**take**	into account the above distinction, but
shall see, studies of digital spaces must	**take**	into account the impact of affordances
distributed when intra-categorical complexity is	**taken**	into account. In this light, I extend the
mother tongue. Since the integrated approach	**takes**	into account all types of poetry,

Take account of is far less frequent than *take into account*. It is used with inanimate subjects, such as *the framework, the toolkit.* And it is mostly used in the active voice, as in the following examples:

Example 54: *take account of* in AEC

semantics of language in the poem. Our toolkit	**takes**	account of the poem's context, both semiotic
historical-discourse analysis. The framework	**takes**	account of the complexity of legitimation

On the contrary, *take into consideration* seems to be used mostly with animate subjects, such as *I,* and seems to prefer the passive voice, which also implies an animate subject.

Example 55: *take into consideration* in AEC

Note, further, that in my analysis I also	**take**	into consideration cases of metaphorical
eight males; the "gender" variable was not	**taken**	into consideration in the processing of

While *take into account* is associated with animate subject, *take account of* is associated with inanimate subjects. In academic writing, *take into account* seems to be preferred. Like *take into account, take into consideration* prefers an animate subject. Things like, *the cases, the impact, the distinction* and *the context* are examples of what is to be considered before making decisions.

TAKE PART IN (13)

In general use, *take part in* activates the semantic frame of 'Participating' which involves participants and an event or institution as core elements. In AEC, the participants are *agents, children, students, participants, voters, they*, all participated in events like *the events, this study, sport,* as in the following examples:

Example 56: *take part in* in AEC

three to five students each–volunteered to	**take**	part in the study, filling out background
discourse when we encourage our children to	**take**	part in and appreciate sport. The current

TAKE ADVANTAGE OF (10)

In general use, *take advantage of,* means make use of something for specific gain. In AEC, this combination was used with animate subjects, such as *writers, students, learners, we,* to mean, as in general use, to utilize something, such as *time, planning, resources, opportunities* for certain benefit.

Example 57: *take advantage of* in AEC

process where writers have the opportunity to	**take**	advantage of more time online for planning
production, where students had the possibility to	**take**	advantage of online planning. In addition

TAKE RESPONSIBILITY FOR (7)

In general use, *take responsibility for something* means to accept the negative consequences of a mistake. This meaning is also found in the AEC, in sentences like the following, where the agent *government, unit,* takes responsibility for the *crisis* and the *budget*.

Example 58: *take responsibility for* in AEC

but also argued that his government cannot	**take**	full responsibility for solving the crisis
salient, because it implies that the unit	**takes**	responsibility for the budget, thereby

However, it is noted that in the AEC, *take responsibility for sth* is not only associated with negative consequences. In the following examples, the agents *learners* (they) and *speakers* took responsibility for their *learning* and *the success of the communication,* which hints at more positive associations.

Example 59: *take responsibiliy for* + positive association in AEC

meaningful to them, they are more likely to	**take**	responsibility for their learning and to
because they are 'devices by which speakers	**take**	responsibility for success in communication

MAKE USE OF (44)

In general use, *make use of* has the same meaning as *take advantage of,* which means to utilize something for certain benefit. In academic writing, *make use of* has the same meaning as that of general use. An animate subject, such as *the participants, teachers, he, they, politicians, the candidates,* or an inanimate subject, such as *this study, posters, Hebrew, English,* utilizes *resources, values, frameworks, strategies, affordances, means* and makes use of them to achieve a certain benefit, as in the following sentences:

Example 60: *make use of* in AEC

shared by colleagues. Teachers regularly	made	use of such resources to develop their
will explore further how the candidates	made	use of various values of modality and linguistic
The corpus component of this study	makes	use of the British National Corpus and
As regards color, institutional posters	make	greater use of color than the other advertisers

MAKE REFERENCE TO (13)

Make reference to, in general use, means to refer to something. It is used similarly in the AEC. Both animate and inanimate subjects could *make reference to something,* as in the following examples,

Example 61: *make reference to* in AEC

deviations from strict ordering typically do not	**make**	reference to the stem. On the other hand
or replication. Nevertheless, we shall	**make**	repeated reference in this article to concepts

Sometimes the word *reference* is fronted in the passive voice indicating that the reference is more important than the person who makes it.

Example 62: *make reference to* in passive voice in AEC

in this frame, explicit references were	**made**	to similar contrasts between the protagonists'
interviewing in Spain, reference must be	**made**	to the legal framework in which the profession

MAKE SENSE OF (11)

In general use, when you make sense of something means that you succeed in understanding it. The same meaning is found in AEC. The subject of *make sense of* is mostly animate, mentioned explicitly as in Example 63. The animate subject succeeds in understanding *past experiences, perspectives.*

Example 63: *make sense of* in AEC

people both inside and outside of government	**make**	sense of each other's perspectives and
Finally, the article considers how the women	**make**	sense of their traumatic experiences, and

DRAW ATTENTION TO (43)

FOCUS ATTENTION ON (12)

TURN ATTENTION TO (8)

The three combinations, *draw attention, focus attention,* and *turn attention,* have *attention* as a fixed noun. The subject of *draw attention* and *focus attention* can be animate or inanimate, as in the following examples

Example 64: *draw attention to* in AEC

synonyms. Thus, teachers should work hard to	**draw**	students' attention to the most frequent
units can be fused into one, Firth (1957)	**drew**	attention to the context-dependent nature
; Leppanen 2012). These studies	**draw**	attention to digital literacy resources
communication. While these developments	**draw**	attention to the meaning relationships

Example 65: *focus our attention on* in AEC

to the meaning of the noncausal verb). We	**focus**	our attention on those kinds of verb pairs
exchange-value. It is in this way that tourism	**focuses**	our attention on linguistic practices as

However, *turn attention to* seems to serve a different function, as it mainly contains a pronoun or a proper noun, such as *we, scholars, I, Kress and Van,* in the subject position and a reflexive pronoun in the object position, the agent *turns his/her attention to something.* This phrase serves a metadiscourse function, as it allows *the speakers, the writer, the agent* to finish one point and guide the reader to the next one. This is further facilitated by the used of the adverb *now*, as in the following examples,

Example 66: *turn attention to* in AEC

multimodal communication, we should now	**turn**	our attention to local practices, the knowledges
studied modes in general, we should now	**turn**	our attention to the specific visual literacies

CALL STH INTO QUESTION (11)

Call into question may be associated, in general use, with the police calling a suspect into question or with causing someone to be viewed with doubt because of certain misbehaviour. However, in the AEC, *assumptions, logic, policies, ideas* are called into question by either animate or inanimate subjects, to be re-examined and evaluated, as in the following examples.

Example 67: *call into question* in AEC

novelty and progress, the structuralists	**called**	into question their assumptions (Carroll
stereotypes in public discourse would otherwise	**call**	into question the educational policies
an anti-categorical approach in that they	**call**	into question the ultimate logic of categories

5.5 Conclusions

This study adopts a usage-based approach to the description of the use of verb-noun collocations within the lexico-grammatical frame of verb complementation clause structures. A contrastive analysis between experts' published academic writing in the field of applied linguistics in two languages, English and Arabic, was conducted with the aim of identifying the similarities and the differences between the use of 15 clause structures and the choice of verbs therein. Highlighting the similarities helps in identifying areas that learners may find easier than areas of discrepancies (Ringbom, 2007). The predictions of the difficulties that Arab learners may encounter based on the differences are verified using learners' authentic data in the following chapter, Study 3.

The frequent use of the transitive and the copular clause structure is a common feature in the AEC and the AAC. The use of prepositional verb type 1 clause structure is more frequent in the AAC, as significantly more verb tokens are used, nevertheless, significantly fewer verb types are used in the AAC, which indicates that prepositional verbs are fewer in the Arabic corpus but they are more frequently re-used. The complex copular clause structure is identified as one of the features of English academic writing explored in this study because significantly more verb types and tokens are used in this clause structure in the AEC.

The analysis based on the measure of faithfulness reveals more details regarding similarities and differences. In the copular clause structure in both the AEC and the AAC, resulting verbs are used, e.g. *become*, */sˁa:ra/* and */ʔsˁbaha/.* However, while *become* is frequently complemented with an adjective, the copular verbs in Arabic that have the meaning of *become* take a nominal complement. This leads to the prediction that Arab Advanced learners may prefer a nominal complement when using the verb *become* in English writing.

In the transitive clause structure, while the use of logical-semantic relationship verbs is similar in both corpora, many differences are identified in the use of reporting

verbs. In the AAC, fewer reporting verbs are found to be used and they are mainly of the 'research type'. This hints at the possibility that Arab advanced learners may be limited to a short list of reporting verbs that may not help them express their critical thinking and evaluations.

As opposed to the common classification of phrasal verbs as features of spoken informal language, this study demonstrates how these verbs are currently used academically. The use of phrasal verbs in academic discourse, often, resembles their use in general language. However, many of the phrasal verbs investigated in this study express less frequent meanings or a different meaning to their meanings in general language use. For example, *set out* is frequently used in academic writing to mean *explain,* which is the least common meaning of this verb in general language. While the verb *work out* means *to plan, to exercise, and to develop* in general language use, when used in academic writing it means mostly *to understand.*

The non-existence of phrasal verbs in Arabic indicates that they form a possible area of difficulty for advanced Saudi learners of English. Furthermore, as suggested in the lexico-grammatical analysis of the phrasal verbs identified in the AAC, some of these verbs have a specific academic meaning and use, such specialized information is not provided for learners in any of the available resources, e.g. dictionaries, which adds to the predicted difficulty of phrasal verbs.

Chapter 6: A contrastive Analysis of verb complementation clause structures and verb-noun collocations in novice writers' corpora

6.1 Introduction

Study 3 presents a corpus-based analysis of the verb complementation clause structures and the verb-noun (VN) collocations embedded within these clause structures, in the writing of advanced Saudi learners of English as compared to native speakers. This analysis adopts the same approach used in Study 2 which utilises the Quirkian Clause structures (Quirk et al., 1985) for the syntactic analysis and Frame Semantics (Fillmore, 2006; Fillmore, 2014, 1982) for the semantic analysis of the verb and its complementation. The study investigates novice advanced Saudi writers' use of 15 clause structures in their authentic writing and compares it to novice native writers' use in order to discover whether advanced Saudi learners tend to overuse, underuse or misuse certain clause structures and VN collocations therein and whether this may be the result of transfer from the first language, Arabic. This chapter starts with a presentation of the results and provides analysis of these results, including a section on learners' errors and lexico-grammatical analysis of those errors. The conclusions are presented in the final section.

6.2 Analysis based on total frequencies

The analysis starts with a comparison of the distribution of the 100 most frequent verbs in the NNC and the NSC over the selected 15 clause structures, presented in Table 6-1. Lists of verbs classified under each clause structure, their raw frequencies, normalised frequencies, and their measure of faithfulness to the clause structures in both corpora are presented in Appendix 7. Table reports of the use of each verb, checklists of clause structures and all coded concordance lines are added to the electronic CD, attached to this thesis.

Table 6-1: Selected clause structures and their frequencies in NNC and NSC

		Examples	Types		LL Value	BIC	Tokens		LL value	BIC
			NNC	NSC			NNC	NSC		
1	**Copular/Linking** SVC (_VP NP/ ADJ)	*She is a teacher The girl seemed restless*	6 (14)	8 (20)	0.45	-10.90	631 (1431)	**698 (1703)**	10.03**	-1.32
2	**Simple Transitive** SVOd (_VP NP)	*Tom caught the ball*	82 (186)	73 (178)	0.07	-11.28	2158 (4895)	**2199 (5365)**	9.17**	-2.18
3	**Phrasal Type 2** SVOd (_VP AdvPart NP)	*Sam picked up the pen Sam picked the pen up*	5 (11)	2 (5)	1.12	-10.23	6 (14)	2 (5)	1.81	-9.54
4	**Prepositional Type 1** SVOp (_VP PP)	*John looked at his watch*	31 (70)	45 (110)	3.72	-7.64	207 (469)	**324 (790)**	35.23 ****	23.88
5	**Phrasal Prepositional Type 1** SVOp (_VP AdvPart PP)	*I look forward to your party*	2 (5)	2 (5)	0.01	-11.35	3 (7)	3 (7)	0.01	-11.34
6	**Complex Copular** SVOC (_VP NP NP/ADJ)	*He considered his uncle a genius/ He found the book relevant*	**23 (52)**	10 (24)	4.36*	-6.99	**183 (415)**	98 (239)	23.58 ****	8.94
7	**Complex Transitive** SVOA (_VP NP PP)	*She put her coat in the hall*	**2 (5)**	0 (0)	2.63	-8.72	**7 (16)**	0 (0)	9.20**	-2.15
8	**Ditransitive/ Double object** (dative, benefactive and depriving clause structures) SVOiOd (_VP NP NP)	*He lent Sam his bike*	5 (11)	5 (12)	0.01	-11.34	50 (113)	65 (159)	3.21	-8.14
9	**Prepositional Type 2a** (alternating) SVOdOp (_VP NP PP)	*He lent his bike to Sam*	2 (5)	3 (7)	0.28	-11.07	8 (18)	13 (32)	1.59	-9.76

		Examples	Types		LL Value	BIC	Tokens		LL value	BIC
			NNC	NSC			NNC	NSC		
10	**Prepositional Type 2b** (non alternating) SVOdOp (_VP NP PP)	*He donated £10 to charity*	2 (5)	4 (10)	0.83	-10.52	2 (5)	**13 (32)**	9.84**	-1.51
11	**Prepositional Type 3** SVOdOp (_VP NP PP)	*I caught sight of him*	4 (9)	4 (10)	0.01	-11.34	14 (32)	10 (24)	0.41	-10.94
12	**Prepositional Type 4a** (animate indirect object) SVOiOp (_VP NP PP)	*They told me about your success*	3 (7)	4 (10)	0.23	-11.13	8 (18)	5 (12)	0.50	-10.85
13	**Prepositional Type 4b** (inanimate indirect object) SVOiOp (_VP NP PP)	*They based the findings on fact*	17 (39)	20 (49)	0.51	-10.84	114 (259)	84 (205)	2.64	-8.71
14	**Phrasal Prepositional Type 2** SVOdOp (_VP NP AdvPart PP)	*They put it down to chance*	2 (5)	0 (0)	2.63	-8.72	**3 (7)**	0 (0)	3.94*	-7.41
15	**Phrasal Prepositional Type 3** SVOiOp (_VP NP AdvPart PP)	*They let me in on the deal*	1 (2)	0 (0)	1.31	-10.04	1 (2)	0 (0)	1.31	-10.04
		Total	187 (426)	180 (440)	0.11	-11.24	3395 (7700)	**3514 (8574)**	19.91 ****	8.56

The distribution of verbs, illustrated in Table 6-1 and in Figures 6-1 and 6-2, reveals some similarities and differences between the two groups of novice writers. The transitive, the copular and the prepositional type 1 clause structures are the most frequently used clause structures by both groups.

The distribution of verb types shows a significant difference in the use of the complex copular clause structure, advanced Saudi learners use far fewer types in this clause

structure (24 in the NSC compared to 52 in the NNC). However, the BIC value indicates that this difference is not robust.

Figure 6-1: Normalised frequency of the 100 verbs (types)in the selected clause structures in NNC &NSC

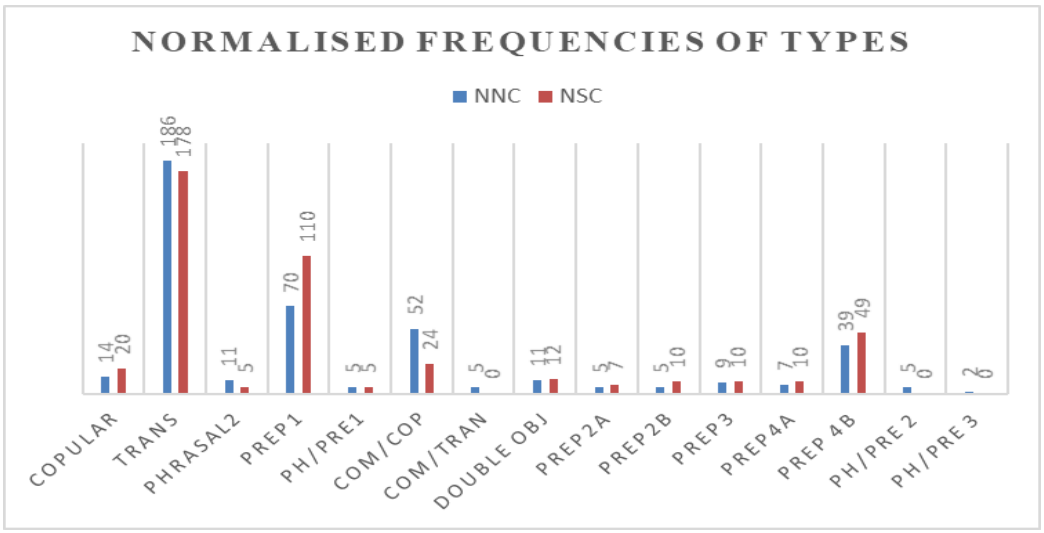
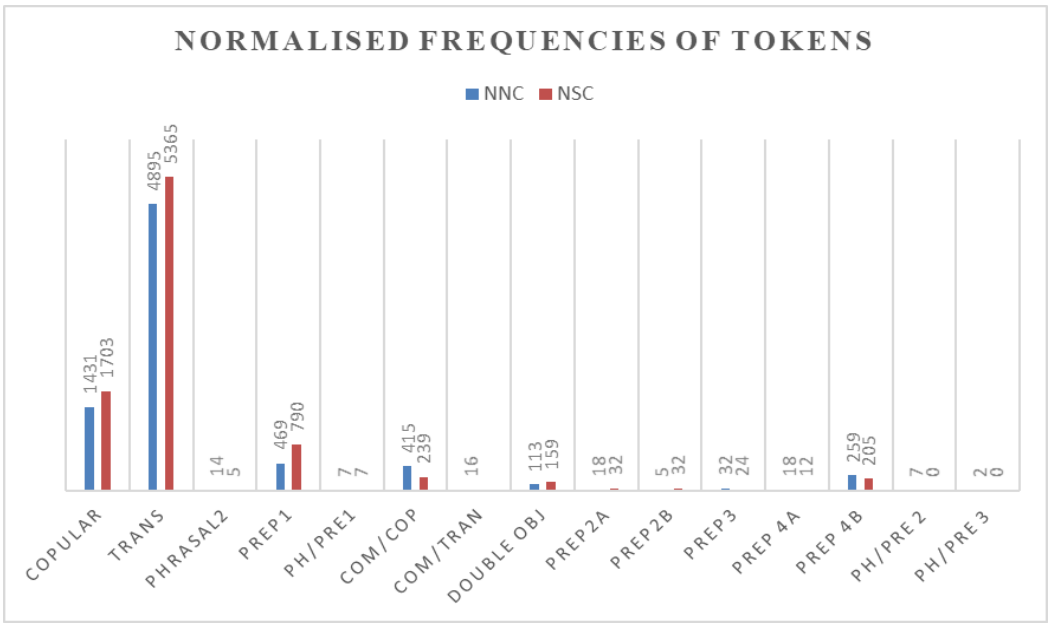


Figure 6-2: Normalised frequency of the 100 verbs (tokens) in the selected clause structures in NNC &NSC



Comparing the distribution of normalised frequencies of verb tokens presented in Figure 6-2 and Table 6-2 shows that there is a robust statistically significant difference in the use of copular, transitive, prepositional type 1 and prepositional type 2b indicating an

overuse by novice advanced Saudi writers as compared to novice native writers. The BIC value undervalues the significant difference in the use of the copular, transitive and prepositional type 2b as it appears to be not worth mentioning. However, the advanced Saudi learners' overuse of prepositional type 1 seems to be very strong with a BIC value of 23.88.

On the other hand, novice native writers use significantly more tokens in the complex copular, complex transitive and phrasal prepositional type 2 clause structures compared to an underuse or absence in the writing of novice advanced Saudi writers. Novice native writers use of significantly more tokens in the complex copular seems to be strong with a BIC value of 8.94.

6.3 Analysis based on the measure of faithfulness

Similar to Study 2, this analysis focuses on clause structures that have more faithful members, specifically verbs that have the measure of faithfulness of 90% or more. By calculating the mean of the measure of faithfulness for each clause structure, presented in Table 6-2, it is observed that the copular and the transitive clause structures have higher levels of faithfulness, especially in the novice native corpus, than other clause structures. This means that many verbs are used in these two clause structures in most of their total occurrences. For example, the verbs *become* and *seem* are used in the copular clause structure in the NNC in 100% of their total occurrences. This indicates that novice writers, both native and Saudi, use verbs in less variety of clause structures and their knowledge may be restricted to the use of these verbs in one or two clause structures.

Table 6-2: Means of the measure of faithfulness of the verbs to the 15 clause structures

		Mean of Faithfulness	
		NNC	**NSC**
1	**Copular/Linking** SVC (_VP NP/ ADJ)	72	37
2	**Simple Transitive** SVOd (_VP NP)	71	64
3	**Phrasal Type 2** SVOd (_VP AdvPart NP)	3	4
4	**Prepositional Type 1** SVOp (_VP PP)	28	18
5	**Phrasal Prepositional Type 1** SVOp (_VP AdvPart PP)	3	5
6	**Complex Copular** SVOC (_VP NP NP/ADJ)	18	28
7	**Complex Transitive** SVOA (_VP NP PP)	10	0

		Mean of Faithfulness	
		NNC	**NSC**
8	**Ditransitive/Double object** (dative, benefactive and depriving clause structures) SVOiOd (_VP NP NP)	33	37
9	**Prepositional Type 2a** (alternating) SVOdOp (_VP NP PP)	10	8
10	**Prepositional Type 2b** (non-alternating) SVOdOp (_VP NP PP)	2	7
11	**Prepositional Type 3** SVOdOp (_VP NP PP)	8	11
12	**Prepositional Type 4a** (animate indirect object) SVOiOp (_VP NP PP)	10	4
13	**Prepositional Type 4b** (inanimate indirect object) SVOiOp (_VP NP PP)	22	17
14	**Phrasal Prepositional Type 2** SVOdOp (_VP NP AdvPart PP)	3	0
15	**Phrasal Prepositional Type 3** SVOiOp (_VP NP AdvPart PP)	6	0
The Mean of Means		19.93	16.00

6.3.1 Copular

Three verbs are more than 90% faithful to the copular clause structure in the NNC, the resulting verb *become* (100), and the stative verbs *seem* (100), and *feel* (90) (see Appendix 7). In the NSC, none of the verbs used is 90% or more faithful to this clause structure (see Appendix7). The faithfulness of the verb *become* in the NSC is affected by the erroneous use of this verb in a number of instances, demonstrated later in Section 6.5.

Although none of the verbs is highly faithful to the copular clause structure in the NSC, the results of the comparison of frequencies, presented in section 6.2, shows that the copular clause structure is overused by novice advanced Saudi learners. This may be attributed to the use of the verb *be.* The frequency of occurrence of the verb *be* in the copular clause structure in the NSC is higher than its frequency in the NNC, it has the normalised frequency of 1,580 in the NSC compared to 1250 in the NNC (see Appendix 7). The measure of faithfulness is also larger (65% in the NSC vs. 55% in the NNC) which indicates that advanced Saudi learners tend to use the verb *be* as a main verb more frequently than

native speakers do. This result supports the findings of Hinkel (2002:113-114) who observes that the use of the verb *be* as a main verb in the copular clause structure, which is a characteristic of spoken discourse, is a common feature of non-native speakers writing. This result also demonstrates the "teddy bears" phenomenon which refers to learners' overuse of words/phrases because they are considered safe choices (Hasselgren, 1994; Nesselhauf, 2005).

While the verb *be* is equally complemented with a nominal and adjectival complement in the NNC, it is followed by a nominal complement in more than 60% of its total occurrences in the copular clause structure in the NSC. The use of the copula *be* with a nominal complement in existential *there*-sentences is common in the two corpora of novice writers, as in the following examples:

Example 68: *be* in NNC

or another language, alternatively. There	**are**	three main arguments to the idea of not
the same motivations as non-natives. There	**is**	a strong argument for native vs. non-native

Example 69: *be* in NSC

attempt to reparse it in different way. There	**are**	2 important principles in this model. Which
that we use to learn the language. There	**are**	many types of cognitive strategies like

However, a noticeable use of the copula *be* with a nominal complement in the NSC is to introduce a definition of a term, as in Example 70. The verb *define* is not among the 100 most frequently used verbs in the NSC, it occurs three times only in the whole corpus. It seems that advanced Saudi writers rely on the verb *be* to provide definitions and could potentially benefit from being taught other verbs that serve the purpose of defining subject matters.

Example 70: *be* + definition in NSC

in the metal operation. 4-19 Dysarthria;	**is**	a speech disorder, resulting of damage
their comprehension is intact. Dyslexia;	**is**	troubles in reading despite general / normal
in the language. The audio-lingual method	**is**	the methods of foreign language teaching

6.3.2 Transitive

The transitive clause structure has a high mean of faithfulness in the two corpora. Although the mean of faithfulness to the transitive clause structure is lower in the NSC, the analysis in Section 6.2 shows that advanced Saudi learners overuse the transitive clause structure as compared to native speakers. The statistically significant overuse of the tokens in the transitive clause structure in the NSC is the result of the fact that advanced Saudi learners seem to rely more on certain verbs and use them more frequently. For example, the verb *have* is used more frequently in the NSC (673 in the NSC vs. 431 in the NNC), see Appendix 7. That is because in more than 80% of its total occurrences the verb *have* is used as a main verb by advanced Saudi learners while it is used in only 53% of its total occurrences as a main verb by novice native writers. This result supports the findings of Altenberg and Granger (2001:174) and Sinclair (1991:79) who report that high frequency verbs such as *have, take, make* are often used by learners as delexical verbs, that is they have little meaning in themselves and are used followed by a noun phrase to indicate simply that someone performs an action. Advanced Saudi learners' overuse of some verbs in the transitive clause structure is also in line with the results of Altenberg and Granger (2001), Nesselhauf (2003, 2005), and Hasselgren (1994) which show that learners tend to overuse common high-frequency verbs (the so-called lexical teddy bears). The verbs *use* and *say* are indicative examples; *use* has a normalized frequency of 959 in the NSC compared to 538 in the NNC and *say* 298 in the NSC compared to 43 in the NNC.

Thirty-two verbs in the NNC and twenty-six verbs in the NSC are used in the transitive clause structure in more than 90% of their total occurrences. These verbs, listed in Table 6-3, include the mental verbs *need, want, choose,* and *mean,* the logical semantic relationship verbs *include*, and *affect*, the reporting verb *suggest*, and the activity verbs *produce, use, show* and *marry* which are used similarly in both corpora. The use of the verb *marry*, in the two corpora, is more topic related to the field of Applied linguistics as some of the exam papers are written for the module *sociolinguistics*, issues related to bilingualism and code switching are discussed with reference to marriage between couples who speak different languages.

The verbs *claim, reflect, find, believe, and identify* (underlined in Table 6-3) are used in both corpora in the transitive clause structure. However, they are used in this clause structure by advanced Saudi learners in more than 90% of their total occurrences, which is

a much higher percentage when compared to their counterparts in the NNC (see Table 6-4). Advanced Saudi learners tend to use these verbs mainly in the transitive clause structure while native speakers use them in other clause structures, particularly the complex copular clause structure. This indicates that advanced Saudi learners are less familiar with the use of these verbs in the complex copular clause structure.

Table 6-3: Semantic classes of the most faithful verbs to the transitive clause structure in the NNC & the NSC

Semantic Class	NNC only	NSC and NNC	NSC only
Activity verbs	*attend, create, demonstrate, highlight, hold, measure, show*	*marry, produce, use*	*activate, generate, pronounce, try*
Reporting verbs	*argue, discuss, explain, express, note, propose*	*Suggest*	*claim, indicate, say,*
Mental verbs	*examine, explore, mean, study, tend, understand*	*choose, need, want*	*acquire, believe, determine, discover, find, hear, identify, process, reflect,*
Logical semantic-relationship verbs	*alter, contain, represent, require*	*affect, include*	*attract, support*

Table 6-4: Some verbs with high faithfulness to the transitive clause structure in NSC and their frequencies in NNC & NSC

		The NNC			The NSC		
	Verb	**Raw**	**Norm**	**Faith%**	**Raw**	**Norm**	**Faith%**
1	*claim*	20	45	83	25	61	100
2	*reflect*	24	54	71	16	39	100
3	*find*	75	170	82	22	54	92
4	*believe*	29	66	74	11	27	92
5	*identify*	15	34	65	20	49	91

The reporting verbs *argue, discuss, explain, express, note,* and *propose*, that are used in the transitive clause structure in the NNC in more than 90% of their total occurrences, are identified among the most frequent reporting verbs in academic writing by Hinkel

(2004:186). However, these reporting verbs are not among the 100 most frequent verbs in the NSC. Native speakers use the reporting verbs *discuss* and *explain* in introductory sentences with an agent (usually *this essay*) that discusses/explains a topic, e.g. *This essay will discuss/ explain,* as in Examples 71 and 72. The mental verb *explore* is also used in a similar manner, e.g. *this essay explores,* Example 73. To achieve the same purpose, advanced Saudi learners use a more colloquial, typical of spoken register, verb combination *talk about,* e.g. *the research talked about, To answer this question we must talk about the theory of modularity.* Advanced Saudi learners' use of a two-word verb in place of a single-word may be explained by the so-called "waffle phenomenon" (Edmondson & House, 1991) which refers to learners' use of more words as compared to native speakers.

It may be useful to note that while the verb *discuss* is mostly complemented with a nominal complement, such as *the ways, the factors*, the verbs *explain* and *explore* seem to prefer a finite clause complement, usually a *wh-clause,* as in the following examples:

Example 71: *discuss* in NNC

of multivalency. The following essay will	**discuss**	these factors after first defining the term of
Ireland, USA and England) 3-5 This essay will	**discuss**	the ways sociolinguistics can study language

Example 72: *explain* in NNC

This approach also helps to	**explain**	***why*** particular attitudes exist in society and
communicative goals hold by each. Crystal	**explains**	***how*** the main reasons for the out break of English

Example 73: *explore* in NNC

multi- word verb. 4. This ***essay*** will	**explore**	how a linguistic feature is 'productive',
and events (Barton, 2002). This ***essay*** will	**explore**	what literacy is as a socially situated

The verbs *propose, argue, express* and *note* and the activity verb *highlight* are used with an agent (usually a person) who communicates a message in sentences like:

Example 74: *reporting verbs* + message in NNC

Gropen et al (1989)	**proposes**	a semantic approach
Quirk (1990)	**argues**	that non-natives do not teach 'standard
Plag (2003)	**notes**	that dictionaries are not always reliable
Jenkins (2010)	**expresses**	many negative responses to the works of Kachru
Bauer (2001)	**highlights**	that although forms such as punishment

To achieve the same purpose (expressing that an agent communicates a message), it seems that advanced Saudi learners are limited to one reporting verb *say* as in Example 75.

Example 75: *say* in NSC

The first group (serial model). This group	**said**	that we finish sentence processing first
In this model they	**say**	that we use all the stages at the same time and
Garden Bath	**says**	that when we read sentence we can get meaning

As far as the object complementation is concerned, the reporting verbs in the NNC take both a nominal complement and a finite clause complement. While the verbs *propose* and *express* seem to prefer a nominal complement, the verbs *argue, note* and *highlight* prefer a finite clause complementation. In the NSC, the verb *say* seems to be frequently used with a finite clause. When followed by a nominal complement, it usually refers to the literal pronunciation of a certain word, as in Example 76.

Example 76: *say (pronounce)* in NSC

Like when he speaks and then stop and	**say**	(numnn) to try to remember and also when
in English it must has arms but if we	**say**	armchair in French it must not has an arm
n't pronounce the "g" in [ing] word, they	**say**	swimming, on the other hand, when women

Another important semantic class in the transitive clause structure that seems to be underused by advanced Saudi learners is the group of logical-semantic relationship verbs. These verbs are used to explain relationships between actions and events by providing causes and/or proofs, such as the verbs *cause, change, combine, follow.* These verbs are important for academic writing and are used to express advanced lexical functions (Hinkel, 2004). Although advanced Saudi learners used the verbs *attract* (27), *affect* (27), *include* (56), *support* (34) with a 100% faithfulness to the transitive clause structure, these verbs are not very frequently used. The verb *include* is used by advanced Saudi learners, but less frequently compared to the NNC in which it is used in (122 normalised occurrences). Advanced Saudi learners use of the verb *include* is mainly in relation to the subject *theory*, which includes *parts of speech,* as in example 77.

Example 77: *include* in the NSC

found in real word so the ***theory*** is not	**include**	this. The solution is distinguishing two
points of weakness are the ***theory*** of naming	**include**	the noun and nominal. That is difficult to
of weakness: we can't extend this theory to	**include**	any other parts of speech but nouns

In the NNC, the verb *include* is mostly used with a noun phrase in the object position, with few examples of a finite-clause. In the subject position, the verb *include* is frequently used in relation to *examples,* as in Example 78. Novice advanced Saudi learners very frequently use the word *example* in combination with the preposition *for,* as in Example 79. The use of the verb *include* to introduce examples seems to be unfamiliar for Saudi learners.

Example 78: *examples* + *include* in NNC

mountain. ***Examples*** of accomplishments	**include**	recover from illness, Kim wrote to the
'within'. ***Examples*** of Endocentric Compounds	**include**	; Schoolboy, handcream, brainsurgeon.
a guideline. ***Examples*** of an alphabetic script	**include**	Roman and Greek. The Roman script is used the

Example 79: *example* in NSC

their vocabulary and their pronunciation. ***For***	**example**	, variation of vocabulary as in Britian (sweet)
at home and used in informal context. ***For***	**example**	, the pronunciation of the vernacular [in] vs.

The logical-semantic relationship verb *represent* is found in 100% of its total occurrences in the transitive clause structure in the NNC with a relative high frequency of over 50 but is not frequently used by advanced Saudi learners. The verb *represent* is more commonly complemented with a noun phrase (NP) than with a finite clause, as in Example 80.

Example 80: *represent* in NNC

a word or an idea. In Chinese, sinagrams can	**represent**	meaning or phonetics or both. A problem
may add additional letters in order to	**represent**	their etymology, for instance Mongolian has a
as a: set of visible or tactile signs used to	**represent**	language in a systematic way with a purpose of

6.3.3 Prepositional type 1

In prepositional type 1 clause structure, three verb types are used in more than 90% of their occurrences in the NNC. These verb types are *refer to* (100%), *depend on* (96), *focus on* (92). In the NSC, two verb types are used with a high level of faithfulness, *belong to* (100%) and *refer to* (95). Although the verbs *depend on* and *focus on* are used in the NSC, their level of faithfulness is lower than 90% due to the erroneous use of a different preposition, **focus in* and **depend in* in the examples: **It focus in visuals and speaking tasks, *Our words choices depend in different things.*

6.3.4 Complex copular

In the complex copular clause structure in both corpora, most verbs have lower levels of faithfulness than 90%, except for the verb *consider* in the NSC. It is used solely in this clause structure (100%) by advanced Saudi learners whereas, in the NNC, *consider* is almost equally used in both the transitive (44%) and the complex transitive clause structure (48%), as in Examples 81 and 82.

Example 81: *consider* in the transitive in NNC

telic and atelic. Finally, this essay will	**consider**	multivalency (Briton, 1988), and how this
Ansen, 1998). Therefore, it is paramount to	**consider**	all three aspects when looking at an affix's

Example 82: *consider* in the complex transitive in NNC

were part of a speech study. This method is	**considered**	unethical, however, as people must be aware
or expanding circles and they cannot be	**considered**	native unless they are from the country

It seems that advanced Saudi learners are not familiar with the possible use of the verb *consider* in the transitive clause structure with the meaning of 'think about' which involves a cognizer and a topic. Learners seem to rely more on the use of *consider* as a categorization verb with the meaning of 'regard' which involves a cognizer, an item and a category, as in Example 83.

Example 83: *consider* in NSC

people in the UK it's the opposite, they	**consider**	it nonprestigious and it's for low class
and his own language in that context is	**considered**	the standard language, that is why she
the standard form [last]. Sociolinguistic	**consider**	the using of vernacular form as a sharp

6.4 Lexico-grammatical analysis of selected clause structures

A number of clause structures do not exist in Saudi leaners' first language, Arabic, where in English the verb and the following particle behave as a single verb; this is not possible in Arabic. These clause structures include phrasal type 2, phrasal prepositional type 1, 2, and 3. In the AEC investigated in Study 2, these clause structures, as well as the prepositional type 3 clause structure which has an idiomatic collocational nature, have a high frequency which called for a detailed lexico-grammatical analysis of some issues related to the semantic roles of the arguments of the verbs and comparisons of literal vs. figurative use and general vs. academic use.

Novice writers, both native and Saudi, use a smaller number of phrasal verbs in phrasal type 2 clause structure. Only five verb types are detected in the NNC and 2 of them are also found in the NSC, which are *find out* and *write down.* In phrasal prepositional type 1, two verb types are found in each corpus. The phrasal prepositional verb *find out about* is shared between the two corpora. *Look out for* is used in the NNC and *come up with* in the NSC. No examples of phrasal prepositional type 2 and 3 clause structures are found in the NSC for the verbs investigated, however, in the novice native writers' corpus, two verb types are found for the phrasal prepositional type 2 clause structure, *made up of* and *add on to* and only one verb type is found for phrasal prepositional type 3 which is *lead on to*. For the prepositional type 3 clause structure, four verb types are found in the NNC, *take sth into account, take sth into considerations, take part in,* and *take control of.* The same number of types is found in the NSC in the prepositional type 3 clause structure but different examples, including *take care of, switch attention to, give attention to* and *make use of.*

The infrequent use of phrasal verb clause structures by novice advanced Saudi learners may be an expected result in this study due to the absence of these clause structures in the learners' first language. This result confirms the results of previous research on learners' use of verb-noun lexical collocations which indicate that learners have greater difficulty with non-congruent collocation, i.e. have no equivalent in the first language (e.g. (e.g. Al-Zahrani, 1998; Nesselhauf, 2003). It is also in line with the results of Dagut and Laufer (1985), Liao and Fukuya (2004) and Waibel (2007) which indicate that the nonexistence of phrasal verbs in learners' first language leads to learners' underuse and/or avoidance of these verbs. However, novice native writers infrequent use of phrasal verbs clause structures comes up as an unexpected result in this study. Given the fact that phrasal

verbs are identified as a characteristic of conversational and informal registers (Siyanova & Schmitt, 2007) which native novice writers are well familiar with from their everyday communication, it was expected that native novice writers are more likely to use phrasal verbs in their writing. It was envisaged that they would not have the difficulty that advanced Saudi learners have in understanding the meaning and the appropriate use of these verb combinations. The low frequency of phrasal verbs in the writing of novice native students seems to indicate that they intend to sound more formal and avoid using any of the characteristics commonly identified as features of spoken and informal register. Although this study cannot confirm the extent to which novice writers, both native and non-native, consciously avoid phrasal and phrasal prepositional verbs, the near absence of these types of verbs in their academic writing might be an effect of direct instructions or guidelines either received from their teachers or textbooks. This could be an interesting area to investigate further, possibly through interviews or other quantitative/qualitative techniques.

Because the frequencies of types in these clause structures are so low, sometimes zero in the NSC, in the novice corpora, it is difficult to perform a lexico-grammatical analysis similar to the one in Study 2. Some of the verb types that occur in these clause structures are found only once in the corpora at hand which increases the possibility of idiosyncratic use rather than a pattern of use on which observations and conclusions can be made.

6.5 Lexico-grammatical analysis of learners' errors

The errors identified in the advanced Saudi learners' corpus, presented in appendix 8, can be classified into four groups. The largest group of errors is related to the use of the preposition in prepositional type 1 clause structure, either the preposition is missing, used redundantly or misused. The verbs *accord, listen, refer* and *speak* are used without the preposition *to* when it is needed to make the structure clearer. The verbs *activate* and *denote* are followed by the preposition *to* which is redundant and deleting it would make the sentence sound more natural. Similarly, the verb *marry* is redundantly followed by the preposition *in*, in the example **The men must marry in outside tribe*. The verb *reach* is redundantly followed by *at* in the example **When we reach at the end of the sentence*. A number of verbs are followed by a preposition where another preposition should have been used instead, for example, **people affected with this type* should be *affected by*, **our word*

choice depend in/to different things should be *depends on, *vocabulary distinguish people to highest class and lowest class* should be *distinguish into, *it focuses in visual and speaking tasks* instead of *focuses on,* and **learn with themselves* instead of *learn by themselves.*

The other group of errors is related to the copular clause structure. In many examples, learners misused the verbs *become* (4 examples), and *appear* (3 examples), as shown in Appendix 8. The verb *seem* is also misused, but only in one instance, that is why it is not considered in this analysis. The verb *become* is a copular verb that expresses a resulting attribute of the subject and it is followed by an adjective or a noun phrase. It is used incorrectly in 4 instances, see Example 84. In the first sentence, it is not used to express a resulting attribute of *interference,* but rather how it is classified. It seems that the verb *become* mistakenly is used instead of the verb *come.* In the second sentence, the preposition *as* followed by a noun phrase is used redundantly. The preposition *as* is often used in the complex copular clause structure to add an attribute to the object, e.g. *we considered him as a genius.* Its use in the copular clause structure is uncommon and results in an odd structure. In the third sentence, although it may be acceptable to use the negative word (not) before the verb *become,* its use in this sentence results in an unacceptable structure. The unacceptability of this structure also results from the use of the *to-infinitive* in the subject complement position which is more frequently filled by an adjective or a noun phrase. A similar kind of error is found in sentence four where a verb is used in the subject complement position. In Arabic, the two equivalents of *become, /sˤa:ra/(become)* and */ʔsˤbaha/(become),* can be used as auxiliary verbs followed by a main verb and this may offer an explanation of the learners use of a finite clause and a verb as complements for *become.*

Example 84: *become* in NSC

1. Interference *become under* the issue of transfer
2. This method *become as an opposite* to audiolingual method.
3. The teacher's purpose has not *become to impose* limits and boundaries
4. It will *become relies* on himself

The verb *appear* is also misused in three instances, see Example 85. In the first sentence, it is used in prepositional type 1 clause structure and this usage is very uncommon

for the verb *appear* which is often used intransitively or in the copular clause structure. In the second sentence, the verb *appear* is used in a transitive clause structure which is also uncommon usage of that verb. In the third sentence, the use of the noun phrase *relationship* results in an odd structure. It seems that the verb *appear* is used in this transitive clause structure to mean *show.* This confusion in the use of the verb *appear* could be the result of the learners' first language negative transfer. The Arabic equivalent of *appear* which is */ðˤa hara/* can be used in both the copular and the transitive clause structures and when used transitively it has the meaning of (to show).

Example 85: *appear* in NSC

1. The audiolingual method *appear by* the need to teach the soldiers
2. After the audiolingual method *appear another method*
3. To his principle *appear unfriendly* relationship

A third group of errors is related to the transitive clause structure. Many transitive verbs are misused, but two verbs are particularly more frequently misused, *learn* (9), and *make* (9). That is why a closer investigation of the errors related to these verbs is provided.

For the verb *learn*, the first three of the errors presented in Example 86 are more related to structuring the rest of the sentence, the first example is related to the use of the *to*-infinitive instead of the bare infinitive, the second is related to a missing *in* to make the sentence more acceptable, and the third is caused by the redundant use of the preposition *of*. In the fourth example, the verb *learn* is complemented with a non-finite *-ing* clause, this seems to be an uncommon complementation for the verb *learn* as it prefers a noun phrase or a finite clause with the relative pronoun *that* or question words, such as *how.* In the fifth example, the learner used the preposition *of* to introduce the object of the verb in a kind of prepositional type 1 clause structure, when the use of a direct object without the preposition is more appropriate.

The verb *learn* activates the semantic frame of 'becoming_aware' which involves a 'cognizer', usually a student learning a 'topic' or a 'phenomenon' by certain 'means' or 'instruments' as core elements. It also can involve the degree of learning usually expressed by adverbs such as *hardly,* and *easily*. In sentence 6, the writer used the verb *learn* followed by an adjective which makes the sentence incomplete and consequently can be considered an error. In the last two examples, the verb *learn* is confused with the verb *teach*; it is used

in ditransitive/ double object clause structure with two objects which resulted in an odd structure. The source of this confusion may be the learners' first language. The verb */yaʕlamu/ (learn/ teach)* in Arabic is used to express the two processes of learning and teaching and in the AAC, it is used in the ditransitive/double object clause structure.

Example 86: *learn* in NSC

1. They should to *learn with* people.
2. The interaction with others results *learning language*
3. The child learn many of word from process of naming
4. Experience leads to *learning strengthening* particular connections
5. Child *learns of this words* by the process of naming
6. Student use imagination to *learn fluent*
7. Imagery *learn the students* to use imagination
8. Learn him how to read.

The verb *make* is also misused in 9 occurrences, see Example 88. *Make* is one of the high frequency verbs in English which, according to Altenberg and Granger (2001), tends to be problematic for learners. That is because the verb *make*, just like other high frequency verbs, such as *give, have,* and *go,* are characterised by high polysemy and they dominate many semantic fields. They are also involved in collocations and idiomatic uses.

In the first five sentences, the errors are related to other parts of the sentence. In the first two sentences the verb *avoid* is used after *make* where using the verb *avoid* by itself without *make* would make the sentence more acceptable. In the third sentence, the article *an* is missing, and the use of a different verb like *run* might be better. The fourth error is related to the use of the collocation "*low structure*" where another collocation, probably "*weak structure*" would be more appropriate. In the fifth sentence, the intended meaning is not clear, it could be "*a test of the right word that was made to the subjects*".

In sentences 6 and 7, the verb *make* is used in the transitive clause structure and is followed by inappropriate word class. In 6, the use of the adjective *wrong* is the cause of the error as a noun phrase should have been used instead. In 7, the use of the verb *moving* instead of the noun phrase, *movement,* resulted in an odd structure.

In sentence 8, *make* is used in the complex copular clause structure where the object complement may be an adjective or a noun phrase but not an adverb. Finally, in 9, the phrasal verb *make off* is used to mean *turn off* or *go over them,* whereas in English *make off* means to steal as in Example 87 from the BNC:

Example 87: *make off (steal)* in BNC

Or are you afraid I'll	**make off**	**with the family jewels?**

Example 88: *make* in NSC

1. Avoidance is want to *make avoid* the subject.
2. If they feel a point is not important, they *make it avoid*
3. They *make experiment* about question
4. They use "tend out" which *make low structure*
5. A test that was *made to subject* to the right word
6. He goes back to the passage because he *make wrong.*
7. We *make automatically moving* to alternative interpretation.
8. Make us more carefully
9. Come to clause boundaries and *make it off*

In addition to the errors presented in appendix 8, a total of 511 errors were identified as grammatical (structural) errors. In most of these 511 instances, learners made mistakes in subject-verb agreement. They usually missed the *–s* for the third person singular. In some instances, the errors occurred when formulating the passive form of the verb; an active voice was used instead of the passive or the auxiliary was missing.

6.6 Conclusions

This study uses an innovative analytical approach to the investigation of verb-noun collocations within the larger context of verb complementation clause structures. Rather than the one-dimensional lexical analysis of this type of collocations, the researcher adopts a lexico-grammatical approach which integrates the semantic and the syntactic information to better understand the use of VN collocations within the frame of verb complementation clause structures in academic writing and the problems that advanced non-native learners of English experience.

At the syntactic level, the results reveal that advanced Saudi learners significantly overuse the copular and the transitive clause structures as compared to native writers. Advanced Saudi learners overuse the verb *be* as a main verb in the copular clause structure, a common feature of non-native speakers writing (Hinkel, 2003). The learners' overuse of the transitive clause structure is due to their overuse of a number of high frequency verbs, such as *have, use,* and *say.* This result supports the findings of Altenberg and Granger

(2001) and Nesselhauf (2003, 2005) which report on learners' overuse of common high-frequency verbs.

The measure of faithfulness provides important implications for the learnability of verbs. As observed by Römer et al. (2015: 56), verbs that are used in more than 90% of their total occurrences in a particular clause structure are more likely to be used in a similar way by learners, such as the transitive verbs *choose, mean,* and *suggest*. On the other hand, verbs that are involved in more than one clause structure, are more likely to be used by learners in only one of them, such as the verb *consider* which is used in two clause structures by native writers but only in the complex copular clause structure by advanced Saudi learners.

Findings based on the measure of faithfulness also support the literature on learners' reliance on a limited set of verbs. Specifically, with relation to reporting verbs, Advanced Saudi learners seem to be limited to fewer verbs that are more associated with colloquial and spoken language, such as *talk about* or *say*. Many academic reporting verbs such as *argue, discuss, explain,* and *express,* that are used by novice native writers, are not used by advanced Saudi learners.

The results of this study show that both groups of novice writers used fewer types and fewer tokens in the phrasal verbs and prepositional type 3 clause structures. An important implication of this underuse is that regardless of their language background novice writers would benefit from being introduced to the use of phrasal verbs in academic writing because they seem appropriate in this context as Study 2 has revealed.

The lexico-grammatical analysis helps in the understanding of learners' errors and providing explanations for some of these errors. Learners incomplete knowledge and understanding of the elements of the clause structure in which the verb is used and the elements of the semantic frame that the verb activates when used in that clause structure resulted in misuse of many verbs. This is obvious in the detailed analysis of the errors related to the verbs *become, appear, learn* and *make.*

Chapter 7: Comparisons across corpora

7.1 Introduction

In this chapter, the results of Study 3 are compared to the results of Study 2 on the basis of three types of analysis: analysis based on total frequencies, analysis based on the measure of faithfulness, and lexico-grammatical analysis of selected clause structures.

7.2 Analysis based on total frequencies

Total frequencies in Study 2, which compares expert writers', in the field of applied linguistics, use of 15 clause structures in two corpora, the AEC and the AAC, reveal some similarities and differences between academic writing in English and Arabic. The copular, the transitive, the complex copular and the prepositional type 1 clause structures are found to be the most frequently used clause structures by both groups of expert writers. Although the frequency of verb tokens in the complex copular show that this clause structure is frequently used in both corpora. English expert writers used significantly more types and more tokens in this clause structure which indicates that this clause structure is one of the features of expert English academic writing. The prepositional type 1 is one of the clause structures that are frequently used in the AEC and the AAC. However, English expert writers use of significantly more types, compared to Arabic expert writers use of significantly more tokens, suggests that Arab expert writers are using fewer prepositional verbs. Study 2 also highlights the use of phrasal verbs as an unexpected feature of current academic writing.

The results of Study 3, which compares novice writers use of the same 15 clause structures investigated in Study 2, reveal that the same four clause structures, namely the transitive, the copular, the prepositional type 1, and the complex copular, are the most frequently used clause structures by both groups. Advanced Saudi writers use significantly more tokens in the transitive, the copular and the prepositional type 1 clause structures. However, in the complex copular clause structure, novice native writers use significantly more types and more tokens. This comes as a predicted result since the complex copular clause structure is less frequently used in the AAC, Study 2.

In line with the results of Study 2, which shows that expert Arab writers in the AAC use significantly more tokens in the prepositional type 1 clause structure, novice advanced Saudi writers also use significantly more tokens in the same clause structure than novice native writers. The significant overuse of the prepositional type 1 clause structure by novice advanced Saudi writers may be therefore viewed as a transfer from the first language, Arabic. In Study 2, expert Arab writers used a considerable number of verb types (44) in this clause structure. Although the frequency of types is significantly less than that of native speakers, which is (75), the total frequency of tokens used in this clause structure by expert Arab writers reflects the significance of this type of verbs in academic Arabic as opposed to academic English, see Table 5-1 in section 5.2. This result adds to the literature on verb-preposition collocations as it contradicts the results of many studies on this type of collocations (e.g. Alsakran, 2011) which reports on its difficulty for Arab learners. However, it is important to note that in Alsakran's study no differentiation was made between prepositional verbs and phrasal verbs. Many phrasal verbs were included in the instruments used for the study, such as *drop off,* and *pick up* besides other prepositional verbs, such as *depend on*, all under verb-preposition collocations. This indicates that literature on verb-preposition collocations needs to be carefully considered as usually the two clause structures, prepositional and phrasal verb clause structures, are classified under one group. This study is significant as it highlights the difference between these two clause structures and shows that prepositional verbs are not very difficult for advanced Saudi learners.

The results show a smaller usage of types and tokens in the complex copular clause structure in the two corpora written by Arabic speakers, both experts and novices. The use of the complex copular clause structure seems to be an important characteristic of English academic writing in the field of applied linguistics. Expert native writers used five times the number of types used by expert Arabic writers and novice native writers used double the number of types used by novice advanced Saudi writers. The verbs used by novice advanced Saudi learners seem to match the verbs used in their first language, Arabic, such as *consider, call,* and *make* all used in the complex copular clause structure with higher frequency and higher measure of faithfulness. These three verbs match the top faithful verbs to the complex copular clause structure in the AAC, */dʒaʕala/(make/ create), /samma/ (call), /yuʕtabaru/ (consider),* and */yuʕadu/(consider/ is no longer/ count).* Other verbs that

come at the top of the list of the most frequent verbs in this clause structure in the NNC, such as *define, find, view,* and *describe* are not used by novice advanced Saudi writers.

7.3 Analysis based on the measure of faithfulness

This analysis focuses on four clause structures where a number of verbs are used in more than 90% of their total occurrences, i.e. they are highly faithful to these clause structures: the copular, the transitive, the prepositional type 1 and the complex copular clause structure.

7.3.1 Copular

The verb *become* is the most frequently used verb in the copular clause structure and it is also highly associated (faithful) with that structure in the AEC (98%) and the NNC (100%). This verb is less frequently used by advanced Saudi learners (only in 10 occurrences) in 71% of its total occurrences. Advanced Saudi learners' underuse of the verb *become* may not be attributed to their first language as two verbs in the AAC with the meaning of *become: /sˁa:ra/(become)* and */ʔsˁbaha/(become)* are used in more than 90% of their total occurrences in the copular clause structure. Furthermore, the predicted difficulty, in Study 2, of an underuse of adjectival complementation for the verb *become* is found to be invalid as advanced Saudi learners used adjectival complement after *become.* However, some instances of advanced Saudi learners' misuse of this verb may be attributed to the negative transfer from their first language. In Arabic, the two equivalents of *become, sˁa:ra/(become)* and */ʔsˁbaha/(become)* can be used as auxiliary verbs followed by a main verb and this may be the cause of the erroneous use of this verb by advanced Saudi learners. Such erroneous use of the verb *become* followed by a main verb is not observed in the use of *become* by Chinese learners in Du (2011) which further offers support for evidence of negative transfer from Arabic.

7.3.2 Transitive

As far as the transitive clause structure is concerned, (for ease of reference) Table 7-1 presents all the verbs that have a high measure of faithfulness to this clause structure - 90% or more - in the three English corpora, the AEC, the NNC and the NSC classified into the four semantic classes. Underlined verbs in the AEC are shared with the NNC and bold verbs are shared with the NSC. Bold verbs in the NNC and NSC are shared between the

two. Table 7-2 presents the semantic classes of the most faithful verbs to the transitive clause structure in the AAC.

It has been noted, in Study 3, that some reporting verbs, that are identified among the most frequent reporting verbs in academic writing (Hinkel, 2004:186), are used by novice native writers but are not among the 100 most frequently used verbs in the NSC. These verbs are *discuss, explain, express* and *note.* These verbs are also among the most frequently used verbs in the transitive clause structure by expert writers in the AEC and they are used in this clause structure in more than 90% of their total occurrences. Arab advanced learners depend on the verb *say* as it is ranked the third most frequent verb in the transitive clause structure in the NSC. The verbs *suggest* and *indicate* are also used by advanced Saudi learners but with lower frequencies, *suggest* (14), *indicate* (11), compared to the verb *say* (122). This could be the result of their first language influence as the verb /*qala/(say)* is the second most frequently used verb in the transitive clause structure in the AAC with the verb */ðakara/(mention)* being ranked the first most frequent verb in this clause structure, both with more than 90% faithfulness to this clause structure.

The result pointing to the underuse of reporting verbs by advanced Saudi learners supports the findings of previous studies on reporting verbs (e.g. Bloch, 2010; Manan & Noor, 2014). The authors argue that the non-native speakers' underuse of reporting verbs is due to the lack of understanding and knowledge of the reporting verbs that satisfy the semantic and syntactic requirements of the sentences.

Table 7-1: Semantic classes of the most faithful verbs in the transitive clause structure in the AEC, the NNC, and the NSC

Semantic Class	AEC	NNC	NSC
Activity verbs	*create, demonstrate, establish, illustrate,* ***produce,*** *provide, reveal, show,*	*attend, create, demonstrate, highlight, hold,* ***marry,*** *measure,* ***produce,*** *show,* ***use,***	*activate, generate,* ***marry, produce,*** *pronounce, try,* ***use,***
Reporting verbs	*address,* ***claim,*** *discuss, explain, express,* ***indicate,*** *note,* ***suggest***	*argue, discuss, explain, express, note, propose,* ***suggest***	*claim, indicate, say,* ***suggest***
Mental verbs	*examine, tend, explore,* ***need,*** *observe,* ***determine****, reflect*	***choose,*** *examine, explore, mean,* ***need,*** *study, tend, understand,* ***want***	*acquire, believe,* ***choose,*** *determine, discover, find, hear, identify,* ***need,*** *process, reflect,* ***want***
Logical semantic-relationship verbs	*contain,* ***include,*** *involve, represent,* ***support,***	***affect,*** *alter, contain,* ***include,*** *represent, require,*	***affect,*** *attract,* ***include,*** *support,*

Table 7-2: Semantic classes of the most faithful verbs in the transitive clause structure in the AAC

Semantic Class	AAC
Activity verbs	*/istaʕmala/ (use), /ħamala/ (carry), /ħaðafat/(omit), /istaxdama/ (use), /ħawala/(try), /yuʃakılu/ (form), /ʕaradˤa/ (show), /tabaʕa/ (follow), /yantˤiqu/ (utter),*
Reporting verbs	*/ðakara/(mention), /qala/(say), /tanawala/ (deal with), /naqala/ (report)*
Mental verbs	*/ʔarada/(want), /qarʔa/ (read), /yufi:du/(benefit), /yadrusu/ (study), /qasˤada/ (intend/ aim), /yaʕni:/ (mean/ care/ pay attention), /χalafa/ (disagree), /samiʕa/ (hear), /yafhamu/ (understand),*
Logical semantic-relationship verbs	*/yaqtadˤi:/(require), /yatatˤalabu/(demand), /yatadˤamnu/ (include), /yuqabilu/ (confront),*

Advanced Saudi learners also seem to underuse verbs from the logical-semantic relationship class, such as the verbs *include* and *represent.* These two verbs have a high measure of faithfulness of more than 90% to the transitive clause structure in the AEC and the NNC, as indicated in Table 7-3. Table 7-3 reports the raw frequency, the normalised frequency in brackets and the measure of faithfulness % for each of the logical semantic verbs in the three corpora.

The verb *include* is used by advanced Saudi learners, but less frequently compared to the other two corpora and Saudi learners seem to be unfamiliar with its use in relation to *examples.* This may be attributed to the learners' first language as the Arabic verbs, that have the meaning of *include*, */yatadˁamnu/ (include)*, */yaʃmalu/ (include)* and */yadˁumu/ (include),* in the AAC, are not used to introduce examples. On the other hand, Advance Saudi learners' underuse of the verb *represent* may not be attributed to their first language Arabic, as the Arabic equivalent */yumaθɪlu/(represent)* is among the ten most frequently used verbs in the AEC in the transitive clause structure.

Table 7-3: Logical-semantic relationship verbs in NSC, NNC & AEC

Verb	NSC	NNC	AEC
attract	11 (27) 100%		
affect	11 (27) 100%	27 (61) 100%	
include	**23 (56) 100%**	**54 (122) 100%**	**975 (99) 98%**
support	14 (34) 100%	10 (23) 71%	285 (29) 96%
alter		16 (36) 100%	
contain		17 (39) 100%	355 (36) 99%
represent		**24 (54) 100%**	**557 (56) 90%**
require		28 (64) 100%	357 (36) 83%
involve		**35 (79) 81%**	**816 (83) 100%**

7.3.3 Prepositional type 1

In this clause structure, the prepositional verb *refer to* comes at the top of the list of the most frequent and most faithful verb types to the prepositional clause structure in all three corpora, the AEC, the NNC, and the NSC. In the Arabic corpus, the AAC, the verb */ʔʃara/ (point to/ mention) ila* (to) which can be seen as the counterpart of *refer to* is also at

the top of the list of the most frequent prepositional verbs with a high measure of faithfulness to this clause structure which indicates an area of positive transfer from the learners' first language. The prepositional verb *refer to* is not listed among the most frequent prepositional verbs used by Spanish learners in Wilcoxon (2014: 29) which further offers support for the effect of the Saudi learners' first language, Arabic, on their frequent use of this prepositional verb.

The prepositional verbs *account for* and *contribute to* are used in the AEC to express logical-semantic relationships with a high frequency and a high measure of faithfulness, however, these two verbs are not used by advanced Saudi learners.

7.3.4 Complex copular

In the complex copular clause structure none of the verbs has a 100% faithfulness in the AEC and NNC. However, in the learner corpus the verb *consider* is used in 100% of its total occurrences in this clause structure.

The verb *consider* is used in the complex copular clause structure by native speakers in less than 50% of its total occurrences, see Table 7-4, because it is also used in the transitive clause structure to mean 'think about' or 'look at', as in Examples 89 and 90. Advanced Saudi learners' use of the verb *consider* in the complex copular clause structure may be influenced by their first language. In the AAC, two verbs were used with the meaning of *consider/ regard* in the complex copular clause structure in more than 70% of their total occurrences, see Table 7-5.

Table 7-4: *consider* in the AEC and the NNC

AEC				**NNC**			
	Raw	**Norm**	**Faith%**		**Raw**	**Norm**	**Faith%**
consider	227	23	35	*consider*	24	54	48

Table 7-5: *consider* in the AAC and the NSC

AAC				**NSC**			
	Raw	**Norm**	**Faith%**		**Raw**	**Norm**	**Faith%**
/yuʕtabaru/ (consider)	130	13	99	*consider*	17	41	100
/yuʕadu/(consider/ is no longer/ count)	289	29	76				

Example 89: *consider* in the transitive clause structure *(look at)* in AEC

of intonation and other embodied conduct.	**consider**	the following extract from a speed date
; Schmidt et al., 1995: 287–288). Let us	**consider**	this example from MATVA, where the use
an omnirelevant feature of interaction.	**consider**	the following example. In this example,
time period covered in the present study,	**consider**	figure 1. Generated from the data in Dalton

Example 90: *consider* in the transitive clause structure *(think about)* in AEC

accomplish. A key aim of the article is to	**consider**	the similarities and differences between
2013). Talbot (2007), for example,	**considered**	the way Jamie Oliver constructed himself
entity 'Katrina' more fully, we also need to	**consider**	the left-hand co-text as well as collocates
for broader meanings. In this section, we	**consider**	key semantic tags in August 2005, September
the corpora as described above. We	**consider**	the categories in the first instance, then

It seems that the learners transfer the use of the verb *consider* in the complex copular from their first language. This transfer may have rendered their learning of the other uses of *consider* in the transitive clause structure which is not found in Arabic.

7.4 Lexico-grammatical analysis of selected clause structures

This section deals with the following clause structures: phrasal type 2, phrasal prepositional type 1, 2, and 3. These clause structures are given particular attention because of many reasons. First, they do not exist in Saudi leaners' first language, Arabic, which makes them more error prone. Other reasons include the reported difficulty of phrasal verbs for second language learners of English, in general and the gap in the literature addressing these types of multi-word verbs. Most research is mainly concerned with listing the most frequent verbs in general language use, little is said about the semantic frames and the semantic roles these verbs involve, nor about their use in academic writing.

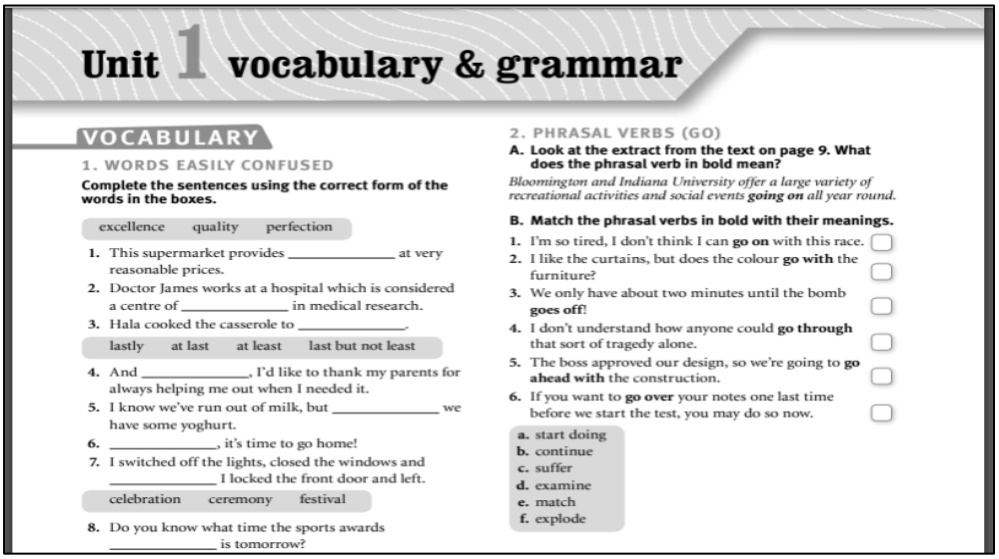
It has been noted in Study 2 that these clause structures have higher frequencies in expert writing than their frequencies in the data of novice writers in study 3. In other words, the results of study 2 and study 3 indicate that expert native English writers tend to use these clause structures more frequently than novice writers, both native and non-native. A total of 70 types of phrasal and phrasal prepositional verb types were used in AEC which constitutes more than 20% of the total of verb types in AEC. Additionally, phrasal verbs

type 2 was the fourth most frequently used clause structure in AEC and phrasal prepositional type 1 was the fifth most frequently used clause structure.

Advanced Saudi learners' limited use of phrasal verbs reported in study 3 confirms the results of previous literature on the use of phrasal verbs by different ESL and EFL learners of English including Hebrew (Dagut & Laufer, 1985), Iranian (Barekat & Baniasady, 2014), Chinese (Liao & Fukuya, 2004), as well as in the Arab world, Omani learners (Abdul Rahman & Abid, 2014), and Egyptian learners (El-Dakhs, 2016). These previous studies attributed learners' underuse of phrasal verbs to a variety of factors, including cross-linguistic differences. Some researchers conclude that the existence of phrasal verbs in learners' first language (such as Germanic languages) facilitates learners' use of these verbs, whereas the absence of phrasal verbs in the first language (i.e. non-Germanic languages such as Hebrew) resulted in learners' underuse and avoidance of these verbs (e.g. Dagut & Laufer, 1985; Liao & Fukuya, 2004; Waibel, 2007). This could provide an explanation for advanced Saudi learners' limited use of phrasal verbs given the fact that in Semitic languages including Arabic, phrasal verbs do not exist.

In addition, textbooks which are designed to teach the English language to Arab students overlook or mention phrasal verbs only in passing (Aldahesh, 2009). In most textbooks used at the school system in the Kingdom of Saudi Arabia (KSA), phrasal verbs are introduced under the vocabulary section and are addressed as part of general spoken language and not as part of academic writing. For example, in the first module of the book (Traveller 6), which is the textbook used for the final year of the high school level, phrasal verbs of the verbs *go* and *turn* are introduced through an exercise in which students are asked to match the phrasal verb with its meaning, Figure 7-1 and 7-2. No explanation is given of their use, their collocations, nor is any reference made to academic use.

Figure 7-1: Phrasal verbs with *go* (Traveller 6)



Unit 1 vocabulary & grammar

VOCABULARY

1. WORDS EASILY CONFUSED

Complete the sentences using the correct form of the words in the boxes.

excellence quality perfection

1. This supermarket provides _____ at very reasonable prices.
2. Doctor James works at a hospital which is considered a centre of _____ in medical research.
3. Hala cooked the casserole to _____.

lastly at last at least last but not least

4. And _____, I'd like to thank my parents for always helping me out when I needed it.
5. I know we've run out of milk, but _____ we have some yoghurt.
6. _____, it's time to go home!
7. I switched off the lights, closed the windows and _____ I locked the front door and left.

celebration ceremony festival

8. Do you know what time the sports awards _____ is tomorrow?

2. PHRASAL VERBS (GO)

A. Look at the extract from the text on page 9. What does the phrasal verb in bold mean?

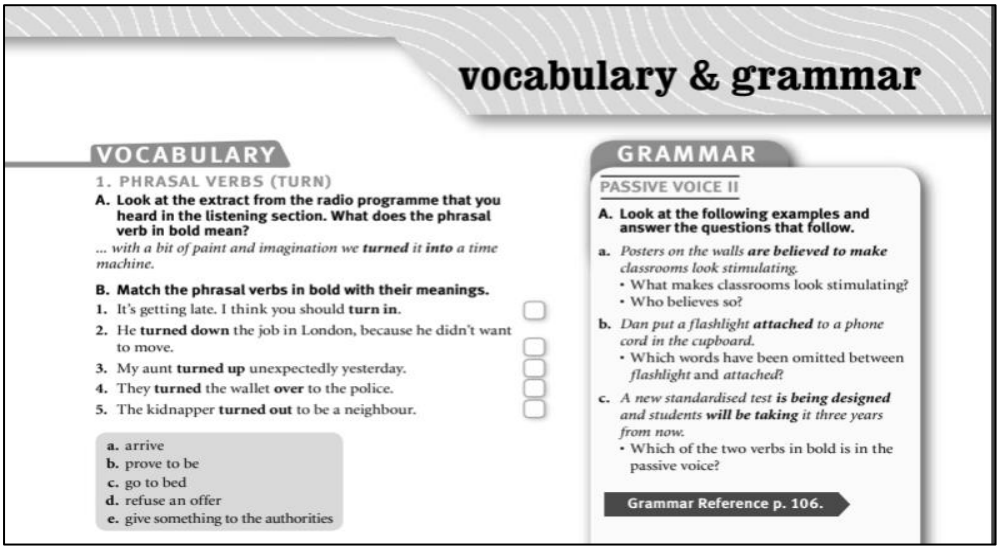
*Bloomington and Indiana University offer a large variety of recreational activities and social events **going on** all year round.*

B. Match the phrasal verbs in bold with their meanings.

1. I'm so tired, I don't think I can **go on** with this race. ☐
2. I like the curtains, but does the colour **go with** the furniture? ☐
3. We only have about two minutes until the bomb **goes off**! ☐
4. I don't understand how anyone could **go through** that sort of tragedy alone. ☐
5. The boss approved our design, so we're going to **go ahead with** the construction. ☐
6. If you want to **go over** your notes one last time before we start the test, you may do so now. ☐

a. start doing
b. continue
c. suffer
d. examine
e. match
f. explode

Figure 7-2: Phrasal verbs with *turn* (Traveller 6)



vocabulary & grammar

VOCABULARY

1. PHRASAL VERBS (TURN)

A. Look at the extract from the radio programme that you heard in the listening section. What does the phrasal verb in bold mean?

*... with a bit of paint and imagination we **turned** it **into** a time machine.*

B. Match the phrasal verbs in bold with their meanings.

1. It's getting late. I think you should **turn in**. ☐
2. He **turned down** the job in London, because he didn't want to move. ☐
3. My aunt **turned up** unexpectedly yesterday. ☐
4. They **turned** the wallet **over** to the police. ☐
5. The kidnapper **turned out** to be a neighbour. ☐

a. arrive
b. prove to be
c. go to bed
d. refuse an offer
e. give something to the authorities

GRAMMAR

PASSIVE VOICE II

A. Look at the following examples and answer the questions that follow.

a. *Posters on the walls **are believed to make** classrooms look stimulating.*
   - What makes classrooms look stimulating?
   - Who believes so?

b. *Dan put a flashlight **attached** to a phone cord in the cupboard.*
   - Which words have been omitted between *flashlight* and *attached*?

c. *A new standardised test **is being designed** and students **will be taking** it three years from now.*
   - Which of the two verbs in bold is in the passive voice?

Grammar Reference p. 106.

Novice native writers' underuse of phrasal verbs as compared to expert writers can be attributed to the fact that the use of phrasal verbs is commonly associated with informal colloquial contexts (Siyanova & Schmitt, 2007). In their study, Siyanova and Schmitt (2007) investigate the use of multi-word verbs as compared to their one-word counterparts in the writing of natives and non-natives. The researchers use the written component of the BNC to investigate native use and the International Corpus of Learner English (ICLE) for learners' use. The researchers note that both native and non-native speakers of English

preferred one-word verbs over their multi-word counterparts. The researchers argued that the participants attempt to select more appropriate words for the register which also conform the expectations of the speech community. This may be one of the reasons behind their preference for one-word verbs over multi-word verbs which are commonly perceived as less formal (p. 121). It is important to note here that Siyanova and Schmitt (2007) used the term multi-word verb as a general term that covers prepositional, phrasal as well as phrasal prepositional verbs, but they intentionally focus on multi-word verbs that are more associated with informal spoken language, such as *put off, go on* and *work out*. This claim is further supported by the results of Chen (2013) who observes that British students use far fewer phrasal verbs than American students. British students' underuse of phrasal verbs may be attributed to the common perception of their informality as highlighted by Siyanova and Schmitt (2007) and thus they may be seen as inappropriate for academic writing.

The high frequency of phrasal verbs in experts' writing in the AEC indicates that the use of phrasal verbs is not restricted to informal registers anymore and is becoming more acceptable in academic writing, at least in the field of applied linguistics. The less frequent use of phrasal verbs in the novice writers' corpus, the NNC, indicates that the association of phrasal verbs and informal contexts persists in teaching contexts and this could be the result of formal instruction in Britain which insists on the informality of phrasal verbs and the need to replace them with a one-word verb when writing academically. For example, the national curriculum in England at the secondary level for the English subject, which was last updated on 2014, (https://www.gov.uk/government/publications/national-curriculum-in-england-secondary-curriculum), specifies among its requirements for the word level at year 6 the understanding of the difference between formal and informal vocabulary and provides one-word verb alternatives for multi-word verbs, see Figure 7-3.

Figure 7-3: Screenshot from the national curriculum in England-English-Appendix2: vocabulary, grammar and punctuation (p.6)

Year 6: Detail of content to be introduced (statutory requirement)	
Word	The difference between vocabulary typical of informal speech and vocabulary appropriate for formal speech and writing [for example, *find out – discover; ask for – request; go in – enter*] How words are related by meaning as synonyms and antonyms [for example, *big, large, little*].
Sentence	Use of the **passive** to affect the presentation of information in a **sentence** [for example, *I broke the window in the greenhouse* versus *The window in the greenhouse was broken (by me)*]. The difference between structures typical of informal speech and structures appropriate for formal speech and writing [for example, the use of question tags: *He's your friend, isn't he?*, or the use of **subjunctive** forms such as *If I were* or *Were they to come* in some very formal writing and speech]

7.5 Conclusions

The results of Study 2 and Study 3 discussed in this chapter lead to many conclusions. In relation to Saudi learners' use of clause structures, advanced Saudis overuse the transitive and the copular clause structure and underuse the complex copular clause structure, which indicates their limited access and familiarity with the complex copular clause structure.

The results of the comparison of the use of the clause structures also indicate that prepositional verbs are not very difficult for advanced Saudi learners. Contrary to what is reported in the literature regarding their difficulty (Alsakran, 2011), advanced Saudi learners, in the NSC, used significantly more types and more tokens in the prepositional type 1 clause structure than native speakers. This result highlights the significance of this study in the literature addressing prepositional verbs as it draws a clear distinction between the two multi-word verb types, phrasal verbs and prepositional verbs, which are often confused in the literature and usually classified together as either prepositional verbs or phrasal verbs. This study specifically indicates that phrasal verbs are underused by learners, but not prepositional verbs. This result may be mainly attributed to the existence of prepositional verbs in the learners first language, Arabic, and the nonexistence of phrasal verbs. The influence of formal teaching and the perceived inappropriateness of phrasal verbs for use in more formal registers might play an important role too.

Phrasal and phrasal prepositional verbs are underused by both native and non-native novice writers. Learners underuse of phrasal verbs supports the literature on the difficulty of this clause structure due to cross-linguistic differences. However, the association of phrasal verbs with informal discourse may be one of the reasons of the underuse of the novice native writers (Dagut & Laufer, 1985; Siyanova & Schmitt, 2007).

Studies on lexical verbs in academic writing tend to focus on one semantic class, with reporting verbs receiving greater attention. However, this study attempts to give an overview of all the semantic classes of verbs used in the transitive clause structure with a special focus on those underused by learners, namely reporting verbs and logical-semantic verbs. Despite their importance in academic discourse, logical-semantic verbs have received little attention in previous literature which has motivated the researcher to give them detailed analysis.

Chapter 8: Conclusions

This chapter presents a summary of the findings of Studies 2 and 3 of this thesis. Afterwards, on the basis of these findings, pedagogical implications are provided. Final conclusions are made at the end of the chapter regarding the contributions, the limitations of the study and suggestions for further research.

8.1 Summary of Findings

The summary of findings is presented in response to the research questions proposed in section 2.8. The first set of questions is related to Study 2 which explored selected verb complementation clause structures and VN collocations embedded within these clause structures in experts' academic writing. The first question is:

1) Which of the selected clause structures are frequently used in expert academic writing in both languages? What are the similarities and the differences in the use of these clause structures, if any, between the two languages?

This question addresses the total frequency of the selected clause structures in the AEC and the AAC which indicates which clause structures are more frequently used by each group of expert writers. The analysis of the total frequencies of verb types in the two corpora reveals that there is no statistically significant difference in the use of verb types in the copular, the transitive and the prepositional type 4a clause structures. This shows that the two groups of expert writers in English and Arabic, use these clause structures similarly with a relatively similar number of verb types. This also indicates that these types of clause structures are prominent features in both academic English and academic Arabic. It can therefore be expected that positive transfer is likely to occur in relation to these clause structures.

Yet, expert native writers use significantly more types in the prepositional type 1, the complex copular, the ditransitive/double object, prepositional type 2a and b, prepositional type 3 and prepositional type 4b. This result indicates a difference between the two languages in the use of these clause structures. It seems that in English academic writing in the field of applied linguistics the use of these clause structures is more common than their use in Arabic in the same domain. The highly significant difference between the two languages in the use of types in the complex copular clause structure (Table 5-1)

highlights this clause structure as a prevalent feature of English academic writing in the studied domain. The ditransitive/ double object and its dative alternation, although possible in Arabic, seem to be disfavoured in academic writing as both have very low frequency of types and tokens in the AAC.

Although expert native writers use more verb types in prepositional type 1, expert Arabic writers use significantly more verb tokens in this clause structure. This result indicates that prepositional verbs are prominent in the academic writing of the two languages in the studied domain. However, English writers use a greater variety of prepositional verbs, while Arabic writers use a more restricted list of types resulting in a higher frequency of tokens.

The second research question is:

2) What are the most frequent verbs used in each of the selected clause structures in both languages? What are the syntactic and the semantic representations of the verbs used in each of the selected clause structures in both languages?

The most frequent verbs used in each clause structure in the two corpora are listed in Appendix 4. In order to address the third question, the measure of faithfulness, which estimates the contingency between the verb and the clause structure, was used to narrow down the analysis to the clause structures that have more faithful members, specifically, verbs that have the measure of faithfulness of 90% or more. Some verbs have a high frequency in a given clause structure but are also frequently used in other clause structures, such as the verb *make* which has the frequency of 519 in the transitive but is also frequently used in the complex copular and the prepositional type 3 clause structures. It is difficult to access this information regarding the occurrence of *make* in these different clause structures depending only on the frequency of occurrence of the verb in the transitive clause structure. Checking the measure of faithfulness of the verb *make* to the transitive clause structure (52%) indicates that this verb is used in this clause structure in 50 percent of its total occurrences. Furthermore, the measure of faithfulness of the verb to the clause structure is useful because clause structures that have more faithful members are easier to learn (Römer et al., 2015: 56), consequently verbs that are used in one clause structure with a high level of faithfulness are easier to learn. On the contrary, verbs that are less faithful, are more versatile and these verbs may pose problems for learners.

Based on the measure of faithfulness, the analysis focuses on four clause structures that have a higher mean of faithfulness than the others. These clause structures are the copular, the transitive, the prepositional type 1 and the complex transitive. For each clause structure the most faithful verbs are analysed in relation to their semantic and syntactic representations.

For the copular clause structure, the resulting verb *become* is very faithful to the copular clause structure in English. Similarly, its Arabic equivalents, in fact two verbs used in Arabic with the meaning of *become /sˤa:ra/* and */ʔsˤbaha/(become),* are also faithful to the copular clause structure. However, while *become* in the AEC prefers an adjectival complement, the Arabic verbs */sˤa:ra/* and */ʔsˤbaha/(become)* prefer a nominal one. This suggests that learners are less likely to use adjectival complements with the verb *become.*

The most faithful verbs to the transitive clause structure in the AEC and the AAC are classified into four semantic classes: activity verbs, mental verbs, reporting verbs and logical-semantic relationship verbs. A similar proportion of activity verbs, mental verbs and logical semantic relationship verbs is used in the AEC and the AAC. However, verbs that belong to the reporting semantic class in the AEC form a longer list than the reporting verbs in the AAC.

The verbs used in prepositional type 1 clause structure with a high level of faithfulness in both languages express a logical-semantic relationship. The verbs *account for, refer to, /dalla/ (indicate) ʕala (on)*, and */ʔʃara/ (point to/ mention) ila (to)* are used to demonstrate the relationship between two constructs of knowledge, such as the relation between *analysis* and *data*, and between a term and its definition.

While in the AEC verbs are not highly faithful to the complex copular clause structure, verbs used in this clause structure in Arabic have higher levels of faithfulness. The verbs */dʒaʕala/(make/ create), /samma/ (call), and /yuʕtabaru/ (consider)* are used in the complex copular clause structure in more than 90% of their total occurrences.

A more detailed lexico-grammatical analysis was carried out for selected clause structures. The decision was made to select the structures that do not exist in Arabic because they are viewed as more error-prone. Prepositional type 3 is included in the analysis because of its idiomatic nature which may impose difficulty on learners. The lexico-grammatical analysis of phrasal verbs underlined the difference between their academic use and the general everyday use. While some phrasal verbs are used in academic writing

exactly in the same senses as in general English (such as *take up* which means *to undertake or handle a role*, both in academic and general use), others are used in selected senses. For example, *work out* is used in academic writing to mean *understand*, a meaning that is used rarely in general language use.

The second set of questions is related to the academic writing of novice native and advanced Saudi writers. The first question in this set is

1) Which of the selected clause structures are frequently used in academic writing by novice writers? Which of these selected clause structures do advanced Saudi learners overuse or underuse as compared to novice native speakers?

In response to this question, the results of Study 3 show that there is no statistically significant difference in the use of types in all clause structures between the two groups of novice writers, except in the complex copular clause structure where advanced Saudi learners show a significant underuse of types and tokens. On the other hand, advanced Saudi learners use significantly more tokens in the copular, the transitive, the prepositional type 1 and prepositional type 2b.

The second research question is:

2) What are the most frequent verbs used in each of the selected clause structure by both groups of novice writers? What are the syntactic and semantic representations of the verbs used in each of the selected clause structures by both groups of novice writers?

As in Study 2, in response to question two, the most frequent verbs used in each clause structure are listed in Appendix 7. The measure of faithfulness is utilised to limit the analysis to the most faithful verbs to the same clause structures explored in Study 2, namely the copular, transitive, prepositional type 1 and complex copular clause structures. The verbs *become, seem* and *feel* are highly associated with the copular clause structure in the writing of novice native writers in the NNC. However, in the NSC, none of the verbs is highly faithful to this clause structure. The most frequently used verb in this clause structure is the copular *be* which is more associated with spoken language.

In the transitive clause structure, advanced Saudi learners rely on a limited set of high frequency verbs, such as *have, use* and *say.* In fact, the verbs *say* and the prepositional verbs *talk about* are the main reporting verbs that learners use to report a message or introduce the topic of the research or the article. Logical-semantic relationship verbs is

another semantic class that is underused among advanced Saudi learners. Verbs like *include* and *represent* which perform important rhetorical functions are underused in the NSC.

The total frequency of types and tokens in the complex copular clause structure shows that advanced Saudi learners use significantly less types and less tokens than novice native writers. The verbs *consider* and *call* are used in this clause structure in more than 90% of their total occurrences in NSC.

Lexico-grammatical analysis based on the semantic frames that phrasal verbs activate could not be performed in Study 3 due to the low and sometimes absence of occurrence of phrasal verbs in the novice corpora. However, lexico-grammatical analysis and semantic frames are very helpful when analysing learners' errors. Semantic frames offer a helpful framework for the investigation of the sources of errors and for tracing the first language influence. For example, the verb *learn* activates the semantic frame of 'Becoming_aware' which involves a 'cognizer', usually a student learning a 'topic' or a 'phenomenon' by certain 'means' or 'instruments' as core elements. It also can involve the degree of learning usually expressed by adverbs such as *hardly,* and *easily*. Learners' error results from the incorrect use of an adjective instead of an adverb after the verb *learn*, **Student use imagination to learn fluent.*

The last question that this study raises is

3) Is there a relationship between advanced Saudi learners' use of clause structures and their first language, Arabic?

The comparison across corpora performed in Chapter 7 shows that while similarities between English and Arabic have a positive impact on the learners' use of the selected clause structures and verb-noun collocations therein, differences between the two languages affect learners' production negatively. Total frequencies of verbs' occurrences in the selected clause structures indicate that the frequent use of verb tokens in the prepositional type 1 clause structure in the AAC may have resulted in Saudi learners' frequent use of this clause structure. On the other hand, the less frequent use of the complex copular clause structure in the AAC may have led to its less frequent use by advanced Saudi learners in the NSC. Furthermore, advanced Saudi learners' limited use of phrasal verbs may be attributed to the nonexistence of this type of multi-word verbs in the Arabic language.

The detailed analysis in Chapter 7 traced the influence of Saudi learners' first language on their use of verbs in different verb complementation clause structures. It seems that the frequent use of a verb in a specific clause structure in the AAC results in a frequent use of the English equivalent of that verb in the NSC. For example, the frequent use of */ʔʃara/ (point to/ mention) ila* (to), the Arabic equivalent of the prepositional verb *refer to*, in the AAC may have resulted in the frequent use of *refer to* in the NSC. The use of */yuʕtabaru/ (consider)* and */yuʕadu/ (consider/ is no longer/ count)*, the Arabic equivalents of the verb *consider*, in the complex copular clause structure in the AAC may have led to learners' frequent use of the verb *consider* in the complex copular clause structure. However, it can be noted that this transfer may affect learners negatively. The fact that the Arabic equivalents of the verb *consider* are mainly used in the complex copular with a high level of faithfulness may have reduced the learnability of the use of *consider* in the transitive clause structure which is observed in the AEC and the NNC. Advanced Saudi learners frequent use of the verb *say*, which reflects the frequent use of the reporting verb */qala/(say)* in the AAC, may have resulted in learners limited use of other reporting verbs in the transitive clause structure.

8.2 Pedagogical Implications

This section starts by presenting an overview of the teaching of English in Saudi Arabia. Then it precedes to present some pedagogical suggestions based on the findings of this study. These pedagogical suggestions are related to two areas: What verbs and verb complementation clause structures to teach? And how?

This study covers 15 clause structures, with a special focus on the copular, the transitive, the prepositional type 1, the complex copular, and phrasal and phrasal prepositional clause structures including phrasal type 1, phrasal prepositional type 1, 2 and 3. Prepositional type 3 clause structure is also touched upon due to the collocability of its elements.

In the Kingdom of Saudi Arabia (KSA), the teaching of English is given a considerable attention based on the view that the English language forms an important means for the progress of the country (Al-Zubeiry, 2012:17-18). The Saudi government allocates a large amount of funding to achieve the goals of English language teaching in the KSA which includes improving students' ability to understand and express themselves

in English (Alharbi, 2015:3). Textbooks dominate the educational process in KSA schools, as is probably the case in most other areas of the world. The Saudi government puts considerable effort into improving the curriculum and the textbooks which results in the introduction of new textbook series from different European and American specialized publishers, such as Macmillan Education. Most of these books focus on teaching English for general use, not for academic purposes.

At the university level, in almost all colleges and universities in the KSA, students have to take a one-year foundation programme with intensive general English language classes (Ahmad, 2012). This course may be supplemented with English for specific purposes which is usually a course designed to provide students with the vocabulary they need to proceed with their studies in their chosen field, such as vocabulary for business, for computer science, or for medicine.

This study addresses advanced Saudi students at university level. It explores their academic writing and investigates their strengths and weaknesses as compared to the academic writing of expert and novice native English speakers. The main purpose of this investigation is to provide useful pedagogical implications and recommendations that may lead to the improvement of students' academic writing in the context of English teaching and learning at the university in KSA. However, it must be acknowledged that choosing expert published articles as the standard for comparison does not mean that the aim is to make all Saudi university students write at the level of published work. It is admitted that not all Saudi students aspire to be expert academic writers or plan to publish their work. Nevertheless, the results and its implications could inform textbook designers and university curriculum planners on the areas that need special focus and reinforcement.

For each of the five main clause structures closely examined in this study, namely, the copular, the transitive, the prepositional type 1, the complex copular, and the phrasal verbs clause structures which include phrasal prepositional verbs as well, a number of recommendations are made.

8.2.1 What verbs and clause structures to teach?

1. Copular verbs

For the copular clause structure, the verb *become* is one of the linking verbs that are more predominant in academic writing than in any other genre (Biber et al., 1999). It is

important to familiarize learners with its use with an adjectival complement as a more preferred complement for *become* than nominal complement.

Advanced Saudi learners seem to be aware of the use of the verb *seem* to perform the function of a hedge that makes the speech less strong and less assertive, as in *this analysis seems to come from the coordination*, instead of *this analysis comes from the coordination*. This highlights an area of strength in the writing of the advanced Saudi learners involved in this study which requires further enhancement. On the other hand, other copular verbs that are frequently used in the AEC, such as *appear* and *remain* are underused or absent in the NSC. Therefore, it is important to familiarize learners with their use in this clause structure.

2. Reporting verbs

Advanced Saudi learners in the NSC seem to be restricted to a limited set of reporting verbs in the transitive clause structure, mainly the verb *say* and *talk about*, both are rather informal conversational options. A longer list of reporting verbs is used by native speakers, such as *discuss, argue, note, explore* and *explain.* It would be very helpful for learners to enrich their vocabulary with these verb choices.

3. Logical semantic relationship verbs

Logical-semantic relationship verbs perform the important function of highlighting the relationship between two constructs of knowledge which is one of the important functions in academic writing (Hinkel, 2004). Advanced Saudi learners underuse some verbs of this semantic class, such as *include* and *represent.* It may be of great benefit to introduce the use of these verbs to learners.

4. The Prepositional verbs

Two prepositional verbs are used in academic expert writing with a high measure of faithfulness, *account for* and *contribute to.* These two prepositional verbs are used to express logical arguments and relations. Their collocational usage may be very helpful for learners when writing academically.

The prepositional verbs *refer to, focus on, depend on* are examples of prepositional verbs that are frequently used across the corpora investigated in this study, including the NSC, with high measure of faithfulness. Which highlights an area of strength in the writing of advanced Saudi learners included in this study.

5. The complex copular

Advanced Saudi learners mainly depend on the verb *consider* in this clause structure, whereas in the AEC and the NNC, more verbs are frequently used in this clause structure, such as *find, view, regard, define, describe.* The use of the complex copular seems to be an important characteristic of English academic writing. It may be helpful to familiarise learners with the use of verbs in this clause structure as this would elevate the level of their writing.

6. Phrasal verbs

As opposed to the long consideration of phrasal verbs as informal and a characteristic of spoken language (Biber et al., 1999), this study highlights the importance of phrasal verbs in current academic English writing in the field of Applied linguistics. More than 20% of the total types used in the AEC are classified as phrasal and phrasal prepositional verbs. Advanced Saudi learners underuse these clause structures. Their use of phrasal verbs is limited to *write down, find out, find out about,* and *come up with*. The frequency of occurrence of these phrasal verbs is less than five which indicates that the use of these verbs may be an idiosyncratic feature of the writing of one or two students.

In the AEC, many phrasal verbs have high frequencies such as *take up* (56), *take on* (32), *take over* (9), *set up* (31), *set out* (13), *make up* (21), *find out* (17), and *work out* (12). Some of these verbs have specific academic use that is different from the general use, such as *work out.* Therefore, it would be very useful to expose learners to these phrasal verbs within the academic context so that they can use them in their own writing.

Many phrasal prepositional verbs are used metaphorically to indicate movement of ideas and thoughts in the 'space' of the article or the research or are used meta-discursively to structure texts; examples include verbs such as: *move away from, come back to,* and *turn away from.* It would be helpful for learners to understand the image of academic writing as a space in which ideas, concepts and beliefs are being moved, shifted etc.

8.2.2 How to teach verbs and clause structures?

This study highlights a number of items that are underused or misused by the sample of advanced Saudi learners investigated in this study. It is hypothesised that exposing the students to these items and explaining their collocations and their use may help improve the academic writing of advanced Saudi learners. The question remains how to introduce learners to the items suggested in Section 8.2.1?

Most textbooks are divided into chapters, each chapter addresses the four skills of listening, speaking, reading and writing, usually, two more sections are added one for vocabulary and the other for grammar. Under the vocabulary section, new vocabulary items are introduced through or followed by an exercise where students are asked to match the word with its meaning or fill in the blank with the correct word. At advanced levels, phrasal verbs and lexical collocations are introduced using the same method. The grammar section focuses on the different tenses including the past, the present, the perfect. In some chapters, the grammar section focuses on reported speech or the formation of the passive voice.

Following Sinclair (1991), Hunston and Francis (2000) and Goldberg (1995), this study adopts the notion that vocabulary and grammar are not separate but are interdependent. As seen from the analysis presented in chapters 4, 5 and 6, in order to understand the meaning of the word, specifically verbs, one must understand the context in which it is used and the collocations it prefers, including grammatical particles which lead to its use as a multi-word verb in a variety of contexts. It is also important to understand the semantic frame that the verb activates, and the semantic roles involved in each semantic frame. Therefore, it is not enough to teach the learner the meaning of the verb under the vocabulary section, and its past, present, future, passive forms under the grammar section. In order to be able to use the verb correctly, learners need a deeper understanding of that verb.

Let us take the verb *find* as an example. In textbooks, it is usually introduced in the vocabulary section as a transitive verb that has many senses (as defined in the Oxford English Dictionary):

1. Discover or perceive by chance or unexpectedly.

   'Lindsey looked up to find Niall watching her'

   'the remains of a headless body had been found'

2. Identify (something) as being present.

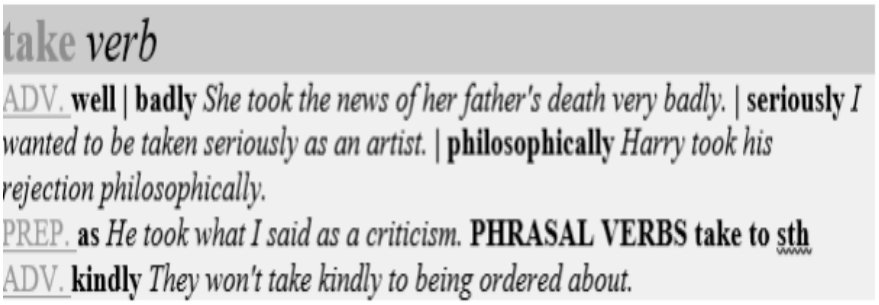
   'vitamin B12 is found in dairy products'

   'a rare species found only in the Italian Alps'

In the grammar section, learners are introduced to the past and the participle form of *find, found,* and how to use this form in sentences like, *yesterday I found my book on the table* and *she has already found her book.*

The dictionary semantic definition of the verb is related to general language use. Such general definition is not always helpful for academic purposes and needs to be expanded to account for the ways in which this verb is used in academic writing. The grammatical description is incomplete as it does not provide any information about the particles that may be used with the verb *find.* Grammatical description is also presented in isolation of the semantic definition. This current study suggests that it would be more useful for learners, who wish to improve their academic writing, to know the clause structures in which the verb *find* is used in academic writing and the semantic frames it activates along with the semantic roles involved. The verb *find,* when used in the transitive clause structure, activates the frame of 'Locating' which involves a perceiver who locates an entity within a ground, such as *this analysis could be found in the online supplementary material.* The verb *find* may also be used in phrasal type 1 clause structure followed by the particle *out*. In this case, it activates the semantic frame of 'Becoming-aware' which involves a cognizer and phenomenon/topic as core elements. The cognizer in the frame of the verb *find out* is mostly an animate. The topic/phenomenon is mostly expressed with a relative pronoun *that, what, etc.* Examples of the use of *find out* include, *the therapist found out what the client had done.* Information about the semantic frames is freely available through the FrameNet database, students can explore by themselves patterns of use of newly introduced verbs.

Dictionaries, specifically online dictionaries, such as the Oxford English Dictionary and Cambridge Dictionary, as well as corpus-based dictionaries of collocations, such as the Oxford Collocations Dictionary for students of English, are useful free available sources for learners. They represent a valuable source for many learners around the world (Nesi, 2012). However, these dictionaries may be largely insufficient. For example, if a learner looks up one of the verbs that are commonly involved in multi-word verb clauses structures, such as *set, make* or *take,* in the Oxford Collocations Dictionary for students of English, he/she will find that the verbs *set* and *make* are not listed at all. *Set* is only represented as a noun and as an adjective, the form *setback* is included but only as a noun and there is no entry for the verb *make,* only an entry for *makeup* as a noun. The entry for the verb *take* includes little information about its adverbial collocations, *well* and *badly,* and some unclear examples of its use with preposition or as a phrasal verb which do not reflect the variety of multi-word combination in which the verb *take* is involved.


take *verb*

ADV. **well | badly** *She took the news of her father's death very badly.* | **seriously** *I wanted to be taken seriously as an artist.* | **philosophically** *Harry took his rejection philosophically.*

PREP. **as** *He took what I said as a criticism.* **PHRASAL VERBS take to sth**

ADV. **kindly** *They won't take kindly to being ordered about.*

Therefore, it is vital for the linguists and language teachers to discuss these dictionaries, criticise them and aim to improve them (Nesi, 2012). Based on the analysis provided in this thesis, the dictionary entry of the verb *take* could be enriched with information in relation to the use of *take* in many multi-word verb clause structures in academic writing which will be of great use for advanced learners of English who aspire to write academically. The improved entry may look as follows:

***take* (verb)**

1. transitive with a following noun phrase.

   Section 4 ***takes*** *a more qualitative approach.*

2. phrasal type 1

   a. *take up* (start to do a particular activity)

      Pre-modifying participles ***take up*** *various roles.*

   b. *take on* (accept a role/ an attribute)

      Compounding may ***take on*** *a number of forms.*

   c. *take over* (gain control)

      This single morph ***takes over*** *the role of the combination.*

3. complex copular

   We ***take*** *these findings to be indicative.*

   These debates are ***taken as*** *a social practice.*

4. prepositional type 3

   a. *take into account/ take account of/ take into consideration*

      She did not ***take into account*** *the above distinction.*

      The toolkit ***takes account of*** *the poem's context.*

      In my analysis, I ***take into consideration*** *the different definitions.*

   b. *take part in*

*This encourages the participants to **take part** in the discussion.*

c. *take advantage of*

*Writers have the opportunity to **take advantage of** the available time.*

d. take responsibility for

*The government cannot **take responsibility for** solving the crisis.*

5. phrasal prepositional type 2 and 3

*This discussion **takes us away from** the main issue.*

*Decontextualizing means **taking something out of** its original context.*

Similar entries can be made for the verbs *set, make* and other verbs based on the classifications and the examples that the data of this study provide which are mainly related to academic writing.

8.3 Contributions

This study makes several theoretical, methodological and pedagogical contributions to research on verb-noun collocations and verb complementation clause structures. The study offers a theoretical review of the representation of the selected clause structures in English and Arabic. This review is particularly useful in the case of Arabic language as many of the clause structures investigated in this study are not labled as such in Arabic grammar books, e.g. the complex copular clause structure. The match between the Quirkian clause structures and the Arabic clause structures attempted in this study represents a comprehensive addition to the literature on Arabic clause structures.

Literature on learners' use of verb-noun collocations mainly adopts the phraseological approach which represents a syntagmatic view of collocations and focuses on lexical items with minimum consideration of grammar. Studies 2 and 3 of this thesis represent a different view of collocations as dependent upon clause structures. This approach provides an understanding of the use of the verb and the following noun phrase at both the syntagmatic and the paradigmatic level. It considers variations associated with the use of the verb with different complements within the same clause structure and across other clause structures. It also considers the variations in the verb slot within the clause structure, that is the association between the verbs and the clause structures.

This thesis is one of the few studies that explores Arab learners authentic writing using corpus tools. Two corpora of Arab learners' writing are specifically compiled for the purpose of this study, a corpus of Arab learners' writing in Study 1 and another corpus of Saudi learners' writing in Study 3. Furthermore, the use of expert corpora, both English and Arabic, is an important contribution of this study to the field of CIA. These two expert corpora represent variations of reference corpora as opposed to the commonly used model of NS/NNS in CIA studies of learners' language. The Arabic corpus is particularly innovative. To the best of the researchers' knowledge, it is the first Arabic corpus specifically designed to include academic journal articles written by expert Arab writers. Available Arabic corpora are either of the Holy book 'Quran' and/or other classical Arabic books, of the writing of learners of Arabic, or of texts randomly selected from the internet, such as arTenTen.

Additionally, the use of the Arabic corpus, the AAC, in this study facilitates the investigation of the influence, positive or negative, of learners' first language, Arabic, on the use of English clause structures. Saudi learners' patterns of use of verbs and verb complementation clause structures is discussed with reference to the patterns identified in the Arabic corpus. This study also addresses the issue of transfer not through investigating lexical similarities between English and Arabic but based on structural similarities between the two languages which can be more useful for learners (Ringbom, 2007:7).

This study provides many pedagogical implications that move away from collocational lists to the lexico-grammatical representation of the use of collocations as embedded within the verb complementation clause structures. This representation offers syntactic and semantic information to support students in contextual usage of collocations in writing.

8.4 Limitations and Suggestions for further studies

This study is limited to academic writing in the field of applied linguistics. A study that includes other disciplines, such as business or law and conducts a comparison across disciplines may be useful to learners of English specialised in other fields.

This study investigates the 100 most frequent verbs in the four corpora. Less frequent verbs are not investigated. A complementary study may explore these less frequent verbs and add information about their use to the findings of this study. Additionally, taking

the most frequent 100 verbs as a starting point for the analysis seems to be a good decision as these verbs represent a good proportion of the data (more than 60% of the total occurrences of all verbs in three of the corpora included, namely the AEC, the NNC and the NSC) which is also practical and less exhausting for the learners. However, this focus on the most frequent 100 verbs rather than the most frequent clause structures is another limitation of this study.

Phrasal verbs are given special attention and are the subject of a detailed analysis in this study. Further research could explore the use of phrasal verbs across other academic discipline to see whether indeed their use is becoming accepted in academic writing overall and what kind of phrasal verbs are seen as appropriate to academic register. Differences between spoken and written academic registers could also be explored. Since phrasal verbs are generally considered features of spoken-ness, it would be interesting to investigate the extent to which phrasal verbs are used in expert academic speech. This could indicate whether there is a tendency of expert academic prose becoming more 'conversational' and similar to academic speech.

The selection of fifteen clause structures for the purpose of this study led to the exclusion of the intransitive clause structure. Many phrasal verbs are used intransitively but are not included in this study as they do not require complementation. A future study that particularly focuses on the use of phrasal verbs in academic writing may include both transitive and intransitive examples of phrasal verbs in order to perform a more comprehensive investigation of the use of these verbs in academic writing.

The near absence of phrasal verbs in novice writers' corpora may be attributed to a variety of factors including direct instructions to avoid the use of phrasal verbs in academic writing and/or the lack of the knowledge and understanding of the most appropriate use of these verbs in academic writing. Further research that explores these factors and maybe discover other possible factors through qualitative or quantitative research techniques, e.g. interviews, would provide practice-based insights into the issue surrounding the teaching of phrasal verbs.

Semantic frames offered a useful framework for the semantic analysis of the selected clause structures in this study. The lexical database, FrameNet, provided information regarding the sense and the semantic frame of the verb as well as the semantic roles associated with that frame. However, the semantic analysis in this study did not turn

out as consistent and comprehensive as intended due to the limitations of FrameNet. Many verb senses are not included in the database. Therefore, in some cases, the semantic analysis of the use of some verbs did not include the semantic frame that the verb activates. Instead, attempts were made to identify the lexical items that filled in the syntactic positions of the subject and the object and the semantic roles that these elements represent.

The results of this study indicate that the ditransitive/double object and its alternation are not commonly used in Arabic academic writing. It would be interesting to examine these two clause structures more closely and explore the influence of their underuse in the AAC on Arabic learners' writing.

This study focuses on the learners' use of VN collocations within the frame of verb complementation clause structures. The main aim was to provide some pedagogical suggestions for the teaching of these lexico-grammatical patterns. It would be good if this study is followed by further research that explores the effectiveness of these pedagogical suggestions.

In this study, analysis of the first language transfer is limited to the discussion of the similarities and the differences between learners' L1 and L2 and learners' L2 performance which mainly focuses on intralinguistic homogeneity. Including data from a group of learners with a different L1 and data of learners' performance in their first language to represent other effects of L1 transfer listed in Jarvis (2000: 253) and Jarvis and Pavlenko (2008: 35), namely intergroup heterogeneity and crosslinguistic performance congruity, would offer more support for the evidence of transfer. However, including these groups was not possible within the time and funding limits of this study. For example, recruiting similar participants in similar settings (e.g. L1 Spanish) would require contacts and visits to universities in other countries, which was not possible within the funding remit for this PhD. This expansion of the data would be an interesting area of future research.

Four different corpora have been compiled for the purpose of this study. These data sets could be used to study other features of academic writing and compare them across the corpora.

References

Abboud, P. F., & McCarus, E. N. (1968). *Elementary Modern Standard Arabic.*: Cambridge University Press.

Abdul Rahman, Z., & Abid, R. (2014). Rarity or non-existence of phrasal verbs in the written discourse of Omani student-teachers of English. *SAGE Open, 4*(4), 1-10.

Ahmad, J. (2012). Theoretical framework & growing demand of ESP in Saudi Arabia. *Arts and Humanities, 65*(5), 114-120.

Ahmed, Z. (2012). *English lexical collocation knowledge of Libyan university students.* (Unpublished doctoral thesis), Bangor University.

Al-Amro, M. (2006). *Saudi learners' knowledge and its relationship to their vocabulary size and writing quality.* (Unpublished doctoral thesis), Colorado State University.

Al-Aqarbeh, R., & Al-Sarayreh, A. (2017). The clause structure of periphrastic tense sentences in Standard Arabic. *Lingua, 185*, 67-80.

Al-Hazmi, S., & Schofield, P. (2007). Enforced revision with checklist and peer feedback in EFL writing: The example of Saudi university students. *Scientific Journal of King Faisal University (Humanities and Management Sciences), 8*(2), 237-267.

Al-Khairy, M. A. (2013). Saudi English-major undergraduates' academic writing problems: A Taif university perspective. *English Language Teaching, 6*(6), 1-12.

Al-Zahrani, M. (1998). *Knowledge of English lexical collocations among male Saudi college students majoring in English at a Saudi university.* (Unpublished doctoral dissertation), Indiana University of Pennsylvania.

Al-Zubeiry, H. Y. A. (2012). The socio-psychological orientations of Saudi learners of English as a foreign language. *Umm Al-Qura University Journal of Languages & Literature, 8*, 13-52.

Aldahesh, A. (2009). *Translating idiomatic English phrasal verbs into Arabic: a contrastive study.* (Unpublished doctoral thesis), University of Western Sydney.

Aldahesh, A. (2013). Teaching English phrasal verbs to non-native speakers of English. *Arab World English Journal, 4*(1), 90-107.

Aldukhayel, D. (2014). *The L2 exposure effect on avoidance of phrasal verbs by Arab ESL learners.* (MA thesis), Retrieved from ProQuest LLC

Alharbi, A. (2015). A descriptive-evaluative study of a Saudi EFL textbook series. *Cogent Education, 2*(1), 1-26.

Alkubaidi, M. (2014). The relationship between Saudi English major university students' writing performance and their learning style and strategy use. *English Language Teaching, 7*(4), 83-95.

Alrabiah, M., Al-Salman, A., & Atwell, E. (2013). *The design and construction of the 50 million words KSUCCA.* Paper presented at the Proceedings of WACL'2 Second Workshop on Arabic Corpus Linguistics.

Alsakran, R. (2011). *The productive and receptive knowledge of collocations by advanced Arabic-speaking ESL/EFL learners.* (Unpublished MA thesis), Colorado State University.

Alshayban, A. (2018). *Avoidance of English phrasal verbs among Saudi ESL students.* (Unpublished doctoral dissertation), The University of Memphis.

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.

Awde, N., & Smith, K. (2004). *Arabic Dictionary: Arabic-English, English-Arabic*. London: Bennett & Bloom.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Barekat, B., & Baniasady, B. (2014). The impact of phrasal verb avoidance on the writing ability of the university EFL learners. *Procedia - Social and Behavioral Sciences, 98*, 343-352.

Barnbrook, G., Mason, O., & Krishnamurthy, R. (2013). *Collocation: applications and implications*. Basingstoke: Palgrave Macmillan.

Bazzaz, F., & Samad, A. (2011). The use of verb noun collocations in writing stories among Iranian EFL learners. *English Language Teaching, 4*(3), 158-163.

Beermann, D. (2001). Verb semantics and double object constructions. *Progress in Grammar: Articles at the 20th Anniversary of the Comparison of Grammatical Models Group in Tilburg.*, Retrieved from https://www.meertens.knaw.nl/books/progressingrammar/beermann.pdf.

Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. Amsterdam: John Benjamins Publishing.

Bhatia, V. (1992). Pragmatics of the use of nominals in academic and professional genres. In L. Bouton & Y. Kachru (Eds.), *Pragmatics and language learning* (Vol. 3, pp. 217-230). Urbana-Champaign: University of Illinois.

Bhatia, V. (1993). *Analysing genre: language use in professional settings*: London: Longman.

Bhatia, V. (2002). A generic view of academic discourse. In J. Flowedew (Ed.), *Academic discourse* (pp. 21-39). Harlow: Longman.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Conrad, S., & Leech, G. (2002). *Longman Student Grammer of Spoken and Written English*. Harlow: Pearson Education.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Biber, D., & Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. In S. Granger (Ed.), *Learner English on computer* (pp. 145-158). London: Longman.

Biskup, D. (1991). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 85-93). Basingstoke: Macmillan Academic and Professional.

Bloch, J. (2010). A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research, 2*(2), 219-244.

Brezina, V., & Gablasova, D. (2017). How to produce vocabulary lists? issues of definition, selection and pedagogical aims. a response to Gabriele Stein. *Applied Linguistics, 38*(5), 764-767.

Charles, M., Hunston, S., & Pecorari, D. (Eds.). (2009). *Academic writing: At the interface of corpus and discourse*. London; New York: Continuum.

Chen, M. (2013). Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics, 18*(3), 418-442.

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(12), 30-49.

Cook, V. (1999). Going beyond the native speaker in language teaching. *Tesol Quarterly, 33*(2), 185-209.

Cowie, A. P. (1998). *Phraseology: Theory, analysis, and applications*. Oxford: Clarendon Press.

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs—A case for contrastive analysis. *Studies in Second Language Acquisition, 7*(1), 73-79.

Daller, M. (2010). Guiraud's index of lexical richness. In: *British Association of Applied Linguistics*, September 2010. Available from: http://eprints.uwe.ac.uk/11902

Danks, W. (2011). *The Arabic verb: form and meaning in the vowel-lengthening patterns* (Vol. 63): John Benjamins Publishing.

Davies, M. (2004). BYU-BNC. *Based on the British National Corpus from Oxford University Press. Available online at http://corpus.byu.edu/bnc*.

De Schryver, G.-M. (2012). Trends in twenty-five years of academic lexicography. *International Journal of Lexicography, 25*(4), 464-506.

Decarrico, J. S. (2001). Vocabulary learning and teaching. *Teaching English as a second or foreign language, 3*, 285-299.

Du, A. (2011). A corpus-based study on use of copular verbs by Chinese school children learning English in China. *i-manager's Journal on English Language Teaching, 1*(4), 33-39.

Edmondson, W., & House, J. (1991). Do learners talk too much? The waffle phenomenon in interlanguage pragmatics. In R. Phillipson, Kellerman, E. Phillipson, R., Kellerman, E., Selinker, L., Sharwood Smith, M., & Swain. M. (Ed.), *Foreign/second language pedagogy research* (pp. 273-286). Clevedon: Multilingual Matters

El-Dakhs, D. A. S. (2016). The lexical knowledge and avoidance of phrasal verbs: The case of Egyptian learners of English. *International Journal of Applied Linguistics and English Literature, 5*(1), 132-144.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *IRAL-International Review of Applied Linguistics in Language Teaching, 33*(4), 315-332.

Farooqui, A. S. (2016). *A corpus-based study of academic-collocation use and patterns in postgraduate Computer Science students' writing.* (Unpublished doctoral thesis), University of Essex.

Faulhaber, S., Herbst, T., & Schmid, H.-J. (Eds.). (2014). *Constructions collocations patterns*. Berlin; Boston: De Gruyter Mouton

Fillmore, C. (2006). Frame semantics. In D. Geeraerts, R. Dirven, & J. Taylor (Eds.), *Cognitive linguistics: Basic readings* (pp. 373-400). Berlin; NewYork: Mouton De Gruyter.

Fillmore, C. (2014). Frames, constructions and frameNet. In S. Faulhaber, T. Herbst, & H.-J. Schmid (Eds.), *Constructions collocations patterns*. Berlin; Boston: De Gruyter Mouton

Fillmore, C. (Ed.) (1982). *Frame Semantics*. Seoul: Hanshin Publishing Company.

Fillmore, C., & Kay, P. (1993). *Construction grammar coursebook: Chapters 1 thru 11*. Berkeley: University of California.

Firth, J. (1957). *Papers in linguistics*. 0xford: Oxford University Press.

Firth, J. R. (Ed.) (1968). *Selected papers of JR Firth, 1952-59*. Harlow: Longmans.

Fries, P. (2002). *Relations and functions within and around language*. London; New York Continuum.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning, 67*(S1), 155-179.

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

Garnier, M. (2016). *English phrasal verbs: usage, knowledge, acquisition.* (PhD thesis), University of Nottingham.

Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research, 19*(6), 645-666.

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41-61.

Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 125-143). Oxford: Clarendon Press.

Gledhill, C. (2011). The 'lexicogrammar'approach to analysing phraseology and collocation in ESP texts. *ASp. la revue du GERAS*(59), 5-23.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago; London: University of Chicago Press.

Goldberg, A. E. (2002). Surface generalizations: An alternative to alternations. *Cognitive Linguistics, 13*(4), 327-356.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Grami, G., & Alzughaibi, M. (2012). L1 transfer among Arab ESL learners: Theoretical framework and practical implications for ESL teaching. *Theory and Practice in Language Studies, 2*(8), 1552.

Grami, G. M. (2010). *the effects of integrating peer feedback into university-level ESL writing curriculum: a comparative study in a Saudi context.* (Unpublished thesis), Newcastle University.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies.* (pp. 37-51). Lund: Lund University Press.

Granger, S. (1998a). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* London Longman.

Granger, S. (1998b). *Learner English on computer*. London: Longman.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (Vol. 6, pp. 3-33). Amsterdam: John Benjamin.

Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly, 37*(3), 538-546.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research, 1*(1), 7-24.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International corpus of learner English. Version 2*. Belgium UCL, Presses Univ. de Louvain.

Granger, S., & Paquot, M. (2009). Lexical verbs in academic discourse: a corpus-driven study of learner use. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 193-214). London; New York: Continuum.

Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on computer* (pp. 119-131). London: Longman.

Green, S., & Manning, C. D. (2010). *Better Arabic parsing: Baselines, evaluations, and analysis.* Paper presented at the Proceedings of the 23rd International Conference on Computational Linguistics.

Greenbaum, S., & Nelson, G. (2002). *An introduction to English grammar* (4 ed.). New York: Routledge.

Greenbaum, S., & Quirk, R. (1990). *A student's grammar of the English language*. Harlow: Longman.

Gries, S., Hampe, B., & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics, 16*(4), 635-676.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective onalternations'. *International journal of corpus linguistics, 9*(1), 97-129.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, 203-257.

Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*: Presses universitaires de France.

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (4 ed.). Gaithersburg: Advanced Analytics, LLC.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23-34.

Halliday, M. (1961). Categories of the theory of grammar. *Word, 17*(3), 241-292.

Halliday, M., & Matthiessen, C. (2004). *An introduction to functional grammar* (3 ed.). London: Arnold.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*(2), 237-258.

Heliel, M. H. (1994). Verb-Particle Combinations in English and Arabic: Problems for Arab Lexicographers and Translators. In R. De Beaugrand, A. Shunnaq, & M. H. Heliel (Eds.), *Discourse and Translation in the West and Middle East* (pp. 141-151). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hilpert, M. (2006). Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory, 2*(2), 243-256.

Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *Tesol Quarterly, 37*(2), 275-301.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Mahwah, NJ: Lawrence Erlbaum Associates.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24-44.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S. (2014). Pattern grammar in context. In S. Faulhaber, T. Herbst, & H.-J. Schmid (Eds.), *Constructions collocations patterns* (pp. 99-119). Berlin; Boston: De Gruyter Mouton

Hunston, S., & Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Hussein, R. F. (1998). Collocations revisited. *Language & Translation (J. of King Saud Univ.), 10*, 39-47.

Hyland, K. (1998). *Hedging in scientific research articles* (Vol. 54). Amsterdam: John Benjamins Publishing.

Hyland, K. (2000). *Disciplinary discourses : social interactions in academic writing*: Harlow : Longman.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*(2), 156-177.

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in them interlanguage lexicon. *Language Learning, 50*(2), 245-309.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*: Routledge.

Javid, C., & Umer, M. (2014). *Saudi EFL learners' writing problems: A move towards solution.* Paper presented at the Proceeding of the Global Summit on Education GSE, Kuala Lumpur, Malaysia.

Jaworska, S., Krummes, C., & Ensslin, A. (2015). Formulaic sequences in native and non-native argumentative writing in German. *International Journal of Corpus Linguistics, 20*(4), 500-525.

Jodai, H., Pirhadi, J., & Taghavi, M. (2014). Attitudes towards native speaker norms: Evidence from an Iranian context. *Procedia-Social and Behavioral Sciences, 98*, 789-798.

Juknevičienė, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. *Kalbotyra, 61*(61), 61-72.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., . . . Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography, 1*(1), 7-36.

Kramsch, C. (2007). Re‐reading Robert Lado, 1957, Linguistics across Cultures. Applied linguistics for language teachers. *International Journal of Applied Linguistics, 17*(2), 241-247.

Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. University of Michigan Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics, 33*(1), 159-174.

Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning, 61*(2), 647-672.

Lee, D. Y. W., & Chen, S. X. (2009). Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing, 18*, 281–296.

Lentzner, K. (1977). *Semantic and syntactic aspects of Arabic prepositions.* (Unpublished doctoral thesis), Georgetown University.

Liao, E.-H. (2010). *An investigation of crosslinguistic transfer in EFL learners' phraseology*. Retrieved from ProQuest LLC.

Liao, Y., & Fukuya, Y. (2004). Avoidance of phrasal verbs: the case of Chinese learners of English. *Language Learning, 54*(2), 193-226.

Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *Tesol Quarterly, 45*(4), 661-688.

Mahmoud, A. (2015). Under-representation of phrasal verbs in EFL university students' free writing. *International Journal of English and Education, 4*(1), 261-268.

Manan, N., & Noor, N. (2014). Analysis of reporting verbs in master's theses. *Procedia - Social and Behavioral Sciences, 134*, 140-145.

Marco, M. (2011). Exploring atypical verb+ noun combinations in learner technical writing. *International Journal of English Studies, 11*(2), 77-95.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

Mel'čuk, I. (1998). Collocations and lexical functions. In A. Cowie (Ed.), *Phraseology. theory, analysis, and applications* (pp. 23-53). Oxford: Clarendon Press.

Namvar, F., Nor, N., Ibrahim, N., & Mustafa, J. (2012). Analysis of collocations in the Iranian postgraduate students' writings. *3L: Southeast Asian Journal of English Language Studies, 18*(1), 11-22.

Nattinger, J. (1988). Some current trends in vocabulary teaching. In R. Carter & M. McCarthry (Eds.), *Vocabulary and language teaching* (pp. 62-82). London: Longman.

Nesi, H. (2012). Alternative e-dictionaries: Uncovering dark practices. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography*. Oxford: Oxford University Press. .

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.

Nesi, H., Sharpling, G., & Ganobcsik-Williams, L. (2004). Student papers across the curriculum: Designing and developing a corpus of British student writing. *Computers and Composition, 21*(4), 439-450.

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

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins Publishing.

Nofal, K. H. (2012). Collocations in English and Arabic: A comparative study. *English language and literature studies, 2*(3), 75-93.

O'Donnell, M., Römer, U., & Ellis, N. (2013). The development of formulaic sequences in first and second language writing. *International Journal of Corpus Linguistics, 18*(1), 83-108.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual review of applied linguistics, 32*, 130-149.

Quirk, R., Greenbaum, S., & Leech, G. (1985). *A comprehensive grammar of the English language*. London: Longman.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 128-143). London: Longman.

Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger (Ed.), *Learner English on computer* (pp. 41-52). London: Longman.

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters.

Römer, U., O'Donnell, M., & Ellis, N. (2014). Second language learner knowledge of verb–argument constructions: Effects of language transfer and typology. *The Modern Language Journal, 98*(4), 952-975.

Römer, U., O'Donnell, M., & Ellis, N. (2015). Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions. In N. Groom, M. Charles, & S. John (Eds.), *Corpora, grammar and discourse: In honour of Susan Hunston (Studies in Corpus Linguistics 73)*. John Benjamins Publishing Company.

Ryding, K. (2005). *A reference grammar of modern standard Arabic*. Cambridge: Cambridge university press.

Ryding, K. (2011). Arabic datives, ditransitives, and the preposition li. *In the shadow of Arabic: The centrality of language to Arabic culture, 63*, 283-299.

Saeed, A. T., & Fareh, S. (2012). Arab EFL learners' acquisition of verbs of senses. *Studia Anglica Posnaniensia, 47*(1), 3-24.

Salih, M. (1985). *Aspects of clause structure in standard Arabic: A study in relational grammar* (Unpublished doctoral dissertation), State University of New York at Buffalo.

Schmid, H.-J. (2014). Lexico-grammatical patterns, pragmatic associations and discourse frequency. In S. Faulhaber, T. Herbst, & H.-J. Schmid (Eds.), *Constructions collocations patterns*. Berlin; Boston: De Gruyter Mouton.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second language research, 12*(1), 40-72.

Shehata, A. K. (2008). *L1 influence on the reception and production of collocations by advanced ESL/EFL Arabic learners of English.* (Unpublished MA thesis), Ohio University.

Siepmann, D. (2005). Collocation, colligation and encoding dictionaries. Part I: Lexicological aspects. *International Journal of Lexicography, 18*(4), 409-443.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (Ed.) (1990). *Collins COBUILD English grammar.* London: Collins.

Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *IRAL - International Review of Applied Linguistics in Language Teaching, 45*(2), 119-139.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review, 64*(3), 429-458.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209-243.

Stein, G. (2017). Some thoughts on the issue of core vocabularies: A response to Vaclav Brezina and Dana Gablasova:'Is there a core general vocabulary?'introducing the new general service list. *Applied Linguistics, 38*(5), 759-763.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford Blackwell.

Subtirelu, N. (2013). What (do) learners want (?): a re-examination of the issue of learner preferences regarding the use of 'native'speaker norms in English language teaching. *Language Awareness, 22*(3), 270-291.

Swales, J. (1981). *Aspects of article introductions*. Birmingham: University of Aston Language studies Unit.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (2004). *Research genres : Explorations and applications*. Cambridge: Cambridge University Press.

Tidball, F., & Treffers-Daller, J. (2007). Exploring measures of vocabulary richness. In M. Daller & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 133 – 149). Cambridge: CUP.

Timmis, I. (2002). Native-speaker norms and International English: a classroom view. *ELT journal, 56*(3), 240-249.

Treffers-Daller, J. (2012). Grammatical collocations and verb-particle constructions in Brussels French: a corpus-linguistic approach to transfer. *International Journal of Bilingualism, 16*(1), 53-82.

Treffers-Daller, J., & Sakel, J. (2012). Why transfer is a key aspect of language use and processing in bilinguals and L2-users. *International Journal of Bilingualism, 16*(1), 3-10.

Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students.* (Unpublished doctoral dissertation), Albert Ludwig University of Freiburg.

Wided, B. (2010). *Modality in English, French and Arabic biomedical siscourse: A contrastive study of drug information leaflets.* (Unpublished MA dissertation), Mentouri University- Constantine.

Wilcoxon, E. M. (2014). *A corpus-based study of the use of prepositional verbs in second language emergent academic writing.* (MA thesis), Retrieved from ProQuest LLC.

Wortabet, J., & Porter, H. (1995). *Hippocrene Standard Dictionary: Arabic-English, English-Arabic*. New York Hippocrene Books.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463-489.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Yanjuan, H. (2014). Analyzing collocation errors in EFL Chinese learners' writings based on corpus. *Higher Education of Social Science, 7*(1), 87-91.

Zareva, A. (2016). Multi-word verbs in student academic presentations. *Journal of English for Academic Purposes, 23*, 83-98.

Zhang, X. (2008). *A contrastive study of reporting in Master's Theses in native Chinese and in Native English.* (Unpublished MA dissertation), The University of Edinburgh.

Appendix 1: Ethical approval for study 1

University of Reading

Memo

School of Literature and Languages
Department of English Language and Applied Linguistics

Please reply to: Dr. Jon Clenton (j.clenton@reading.ac.uk)

To Dr. Sylvia Jaworska

From Dr. Jon Clenton

Copy Manal Alangari
~~Anne Whitbread (file)~~

Date 27th March 2015

Your application for Ethical Approval

Your project entitled **'A corpus study of collocations in the writing of advanced Arab learners of English'** has been considered by the School Ethics Committee, and we are pleased to report that the Committee has raised no ethical objections, and subject to your undertaking to store the consent forms in the Department Office in the normal way, it has accordingly given permission for the project to proceed under the exceptions procedure as outlined in paragraph 6 of the *University's Ethics Guidance to Schools*.

Signed

Dr. Jon Clenton

On behalf of the School Ethics Committee,
Prof. Catherine Leglu, School Director of Research.
Prof. Alison Donnell, Head of School.

Appendix 2: Lists of extracted collocations

1. LOCNESS-A-Level (verb-noun collocations)

	Correct collocations	Tokens	Misused collocations	Tokens
1	takes time	1	fuel the country	1
2	save time	1	secure his family	1
3	waste time	1	menace society	1
4	make the work	1	rising use	2
5	do the work	1	threaten an increase	1
6	continue the work	3	blow the situation	2
7	solve the problem	2	amplify the fact	1
8	face the problem	1	enter the sport	3
9	combat the problem	1		
10	eliminate the problem	1		
11	spend money	7		
12	make money	2		
13	lose money	2		
14	have money	3		
15	give money	3		
16	allocate money	1		
17	waste money	1		
18	put less money	1		
19	invest money	1		
20	get money	2		
21	cost money	1		
22	start life	1		
23	save (his) life	1		
24	ruin (his) life	3		
25	risk (his) life	1		
26	have life	1		
27	run the country	4		
28	decrease number	2		
29	increase number	3		
30	reduce the number	5		
31	growing number	1		
32	start a family	1		
33	create potential world	1		
34	create a society	1		
35	make an example	1		
36	set an example	1		
37	show the effect	1		
38	reduce the effect	1		

	Correct collocations	Tokens	Misused collocations	Tokens
39	have (an) effect	6		
40	give use	2		
41	encourage the use	1		
42	clarify the issue	1		
43	address the issue	1		
44	find (a) way	3		
45	take responsibility	4		
46	have responsibility	1		
47	abdicate responsibility	1		
48	shoulder the responsibility	2		
49	bear the responsibility	2		
50	stop the production	1		
51	reduce the need	2		
52	show the need	1		
53	have (a) need	2		
54	cause an increase	3		
55	take part	6		
56	change the situation	1		
57	take place	20		
58	has a place	1		
59	using technology	1		
60	have (another) reason	1		
61	reduce (public) demand	1		
62	ignore the fact	1		
63	gain power	1		
64	misuse power	1		
65	acquire power	1		
Total number of tokens		134		12

2. TEEP-ArSL (verb-noun collocations)

	Correct collocations	Tokens	Misused collocations	Tokens
1	have time	23	coordinate time	1
2	spend time	38	run the money	1
3	take time	3	afford better life	1
4	save time	2	effect (our) life	2
5	waste time	1	impact the country	1
6	finish work	5	damage the country	1
7	get work	2	effect the country	1
8	provide work	2	decline the number	1
9	do work	4	effect all family	1
10	complete work	1	worries the world	1
11	solve this problem	30	reach the society	1

	Correct collocations	Tokens	Misused collocations	Tokens
12	discuss the problem	4	achieve the society	1
13	reduce the problem	2	effect the society	1
14	face the problem	3	decrease negative effect	1
15	tackle this problem	1	hold (an) effect	1
16	resolve the problem	1	diminish the use	1
17	overcome this problem	1	lack this issue	1
18	have a problem	3	solve the issue	1
19	cause some problem	1	lose their future	1
20	bring hard problem	1	peak their production	1
21	avoid the problem	2	cope the need	1
22	address the problem	1	secure the need	1
23	earn money	14	reduce the situation	1
24	save money	4	increase the situation	1
25	have money	8	stop the situation	1
26	spend more money	1	cope the demand	1
27	get money	1		
28	cost money	1		
29	collect money	1		
30	steal the money	1		
31	lose social life	2		
32	have life	2		
33	reduce the number	4		
34	decrease the number	3		
35	support family	1		
36	face the world	4		
37	make the world	1		
38	lead the world	1		
39	encourage the use	1		
40	discuss the issue	4		
41	tackle the issue	1		
42	highlight this issue	1		
43	face some issue	1		
44	address the issue	1		
45	have a future	1		
46	have a way	1		
47	change the way	1		
48	revolutionised the way	1		
49	have responsibility	4		
50	increase the production	1		
51	reduce the need	1		
52	have a need	1		
53	take part	2		
54	play a part	1		

	Correct collocations	Tokens	Misused collocations	Tokens
55	face the situation	1		
56	discuss the situation	2		
57	take place	4		
58	supply food	1		
59	eat food	1		
60	use technology	3		
61	meet the demand	1		
62	make a case	1		
63	do sport	2		
64	stat my opinion	1		
65	use power	3		
Total number of tokens		223		27

3. LOCNESS-A-Level (Adjective-noun collocations)

	Correct collocations	Tokens	Misused collocations	Tokens
1	strong opinion	1	greater reasons	1
2	nuclear power	2	infected food	1
3	immense power	2	unsavoury places	1
4	popular sports	4	average situation	1
5	dangerous sport	4	related country	1
6	good sport	1	motivated world	2
7	cruel sport	1		
8	barbaric sport	2		
9	public demand	1		
10	low demand	1		
11	big demand	1		
12	new technology	2		
13	modern technology	1		
14	medical technology	1		
15	high echnology	1		
16	current technology	1		
17	main reason	7		
18	good reasons	2		
19	valid reasons	1		
20	plausible reason	1		
21	personal reason	1		
22	various reasons	2		
23	simple reason	1		
24	unhealthy food	1		
25	raw food	1		

	Correct collocations	Tokens	Misused collocations	Tokens
26	processed food	1		
27	fatty foods	1		
28	fast food	2		
29	safer place	1		
30	real situation	1		
31	financial situation	1		
32	other part	1		
33	major part	3		
34	large part	2		
35	various part	1		
36	huge part	1		
37	essential part	1		
38	large increase	2		
39	huge increase	2		
40	marked increase	1		
41	individual need	1		
42	human need	1		
43	moral resposibility	11		
44	ultimate responsibility	1		
45	present day	1		
46	rainy day	1		
47	other days	1		
48	next day	1		
49	early day	1		
50	several way	2		
51	long way	2		
52	efficient way	4		
53	simple way	1		
54	possible way	1		
55	new way	1		
56	inefficient way	1		
57	different way	1		
58	best way	1		
59	near future	3		
60	great future	1		
61	forseable future	1		
62	moral issue	5		
63	major issue	3		
64	local issue	1		
65	important issue	1		

	Correct collocations	Tokens	Misused collocations	Tokens
66	ethical issue	1		
67	difficult issue	1		
68	controversial issue	2		
69	contentious issue	1		
70	widespread use	1		
71	drastic effect	1		
72	beneficial effect	3		
73	significant effect	1		
74	possible effect	1		
75	positive effect	2		
76	overall effect	1		
77	net effect	1		
78	immediate effect	1		
79	ill effect	1		
80	harmful effect	1		
81	devastating effect	1		
82	detrimental effect	1		
83	deleterious effect	1		
84	damaging effect	1		
85	adverse effect	2		
86	big money	4		
87	extra money	1		
88	british government	3		
89	weak government	2		
90	conservative government	3		
91	strong governement	1		
92	czechoslovakean government	1		
93	developed countries	2		
94	western countries	1		
95	underdeveloped country	1		
96	great country	1		
97	other countries	5		
98	royal family	3		
99	small number	5		
100	random number	3		
101	limited number	2		
102	large number	2		
103	whole number	1		
104	sufficient number	1		
105	significant number	1		

	Correct collocations	Tokens	Misused collocations	Tokens
106	odd number	1		
107	large number	1		
108	infinite number	1		
109	huge number	1		
110	equal number	1		
111	third world	4		
112	whole world	1		
113	western world	1		
114	physical world	1		
115	modern world	1		
116	first world	1		
117	developed world	1		
118	modern society	4		
119	bad example	4		
120	obvious example	2		
121	good example	2		
122	simple example	1		
123	classic example	1		
124	strong case	2		
125	extreme case	2		
126	tragic case	1		
127	modern life	2		
128	later life	2		
129	normal life	1		
130	human life	2		
131	healthy life	1		
132	full life	1		
133	everyday life	1		
134	entire life	1		
135	daily life	1		
136	animal life	1		
137	major problem	6		
138	main problem	6		
139	great problem	2		
140	psychological problem	1		
141	financial problem	1		
142	big problem	1		
143	similar problems	1		
144	severe problem	1		
145	fundamental problem	1		

	Correct collocations	Tokens	Misused collocations	Tokens
146	long time	5		
147	short time	3		
148	peak time	3		
149	recent time	2		
150	present time	2		
151	modern time	2		
152	difficult time	1		
153	good work	2		
154	extensive work	1		
Total number of tokens		268		7

4. TEEP-ArSL (Adjective-noun collocations)

	Correct collocations	Tokens	Misused collocations	Tokens
1	solar power	3	last opinion	1
2	high demand	1	respectful places	1
3	modern technology	2	historical place	1
4	new technology	1	huge issue	1
5	main reason	5	passive effect	1
6	important reason	2	large money	1
7	sound reason	1	appropriate money	1
8	good reason	1	outside country	1
9	healthy food	1	east country	2
10	safe place	3	safer countries	1
11	important place	1	mass numbers	2
12	beautiful place	1	several number	1
13	financial situation	1	numerous number	1
14	important part	2	added number	1
15	substantial part	1	individual society	1
16	small part	1	global society	2
17	significant part	1	healthy example	1
18	large part	1	luxury life	1
19	essential part	1	wealthy life	1
20	active part	1	economical problem	3
21	significant increase	4	negative problems	1
22	rapid increase	1	largest problem	1
23	marked increase	1	over-working problems	1
24	gradual increase	1	healthy problem	1
25	dramatic increase	1	hard problem	2
26	considerable increase	1	lesser time	1
27	urgent need	4	higher time	1
28	social need	2		
29	basic need	2		
30	physical need	1		

	Correct collocations	Tokens	Misused collocations	Tokens
31	personal need	1		
32	individual need	1		
33	great need	1		
34	serious responsibility	1		
35	individual responsibility	1		
36	enormous responsibilities	1		
37	global production	1		
38	best way	10		
39	effective way	6		
40	proper way	2		
41	various way	1		
42	traditional way	1		
43	several way	1		
44	new way	1		
45	efficient way	1		
46	alternative way	2		
47	near future	2		
48	forseeable future	2		
49	controversial issue	4		
50	major issue	3		
51	important issue	2		
52	debatable issue	3		
53	social issue	1		
54	political issue	1		
55	main issue	1		
56	global issue	1		
57	environmental issue	1		
58	different use	2		
59	positive effect	12		
60	undesirable effect	2		
61	important effect	2		
62	direct effect	1		
63	detrimental effect	3		
64	adverse effect	2		
65	Saudi governemnt	1		
66	Danish governement	1		
67	poor country	3		
68	European country	5		
69	rich country	4		
70	developing country	2		
71	hot country	1		
72	developed country	1		
73	different countries	5		
74	various countries	1		

	Correct collocations	Tokens	Misused collocations	Tokens
75	poorer country	1		
76	poor family	1		
77	good family	1		
78	unlimited number	1		
79	huge number	6		
80	significant number	6		
81	considerable number	7		
82	high number	3		
83	great number	2		
84	substantial number	1		
85	limited number	2		
86	whole world	2		
87	modern world	2		
88	good example	5		
89	greate example	1		
90	excellent example	1		
91	social life	25		
92	daily life	9		
93	human life	5		
94	healthy life	3		
95	good life	3		
96	personal life	2		
97	normal life	2		
98	work life	1		
99	real life	1		
100	private life	1		
101	modern life	2		
102	happy life	1		
103	full life	1		
104	busy life	1		
105	social problem	6		
106	serious problems	5		
107	main problem	4		
108	psychological problem	3		
109	major problem	2		
110	financial problem	1		
111	environmental problem	1		
112	economic problem	2		
113	complex problem	1		
114	common problem	1		
115	mental problems	2		
116	biggest problem	2		
117	long time	40		
118	short time	7		

	Correct collocations	Tokens	Misused collocations	Tokens
119	free time	6		
120	recent time	2		
121	present time	1		
122	good time	1		
123	spare time	1		
124	semi-skilled work	1		
125	manual work	1		
126	intensive work	1		
127	hard work	2		
128	good work	1		
129	difficult work	1		
130	daily work	1		
Total number of tokens		268		7

Appendix 3a: list of the top 100 verbs in the AEC

	Lempos	Raw Freq	Norm Freq
1	be-v	35,786	3624.31
2	have-v	6,804	689.09
3	do-v	2,828	286.41
4	use-v	2,405	243.57
5	see-v	2,257	228.58
6	show-v	1,457	147.56
7	make-v	1,382	139.97
8	follow-v	1,207	122.24
9	give-v	1,124	113.84
10	include-v	1,086	109.99
11	take-v	1,064	107.76
12	provide-v	1,000	101.28
13	find-v	975	98.75
14	involve-v	819	82.95
15	speak-v	725	73.43
16	come-v	677	68.56
17	suggest-v	668	67.65
18	consider-v	654	66.24
19	argue-v	630	63.80
20	occur-v	616	62.39
21	represent-v	616	62.39
22	discuss-v	614	62.18
23	refer-v	594	60.16
24	become-v	592	59.96
25	mean-v	591	59.85
26	appear-v	586	59.35
27	focus-v	584	59.15
28	seem-v	562	56.92
29	identify-v	562	56.92
30	say-v	554	56.11
31	describe-v	549	55.60
32	base-v	532	53.88
33	express-v	518	52.46
34	indicate-v	515	52.16
35	present-v	508	51.45
36	go-v	501	50.74
37	know-v	499	50.54
38	construct-v	487	49.32
39	allow-v	472	47.80

	Lempos	Raw Freq	Norm Freq
51	write-v	417	42.23
52	report-v	414	41.93
53	relate-v	413	41.83
54	explain-v	403	40.81
55	propose-v	399	40.41
56	associate-v	396	40.11
57	reflect-v	390	39.50
58	develop-v	375	37.98
59	learn-v	362	36.66
60	look-v	360	36.46
61	contain-v	357	36.16
62	define-v	353	35.75
63	assume-v	347	35.14
64	accord-v	338	34.23
65	compare-v	336	34.03
66	work-v	334	33.83
67	establish-v	333	33.73
68	demonstrate-v	330	33.42
69	regard-v	318	32.21
70	offer-v	315	31.90
71	get-v	309	31.29
72	ask-v	307	31.09
73	call-v	307	31.09
74	reveal-v	307	31.09
75	remain-v	306	30.99
76	think-v	300	30.38
77	support-v	297	30.08
78	expect-v	296	29.98
79	observe-v	296	29.98
80	address-v	295	29.88
81	analyse-v	290	29.37
82	play-v	285	28.86
83	set-v	281	28.46
84	help-v	274	27.75
85	turn-v	273	27.65
86	share-v	269	27.24
87	contribute-v	269	27.24
88	start-v	256	25.93
89	begin-v	256	25.93

	Lempos	Raw Freq	Norm Freq
40	need-v	465	47.09
41	examine-v	455	46.08
42	lead-v	437	44.26
43	require-v	432	43.75
44	understand-v	431	43.65
45	produce-v	428	43.35
46	apply-v	426	43.14
47	draw-v	424	42.94
48	illustrate-v	424	42.94
49	note-v	421	42.64
50	create-v	421	42.64

	Lempos	Raw Freq	Norm Freq
90	claim-v	255	25.83
91	introduce-v	254	25.72
92	mark-v	253	25.62
93	increase-v	248	25.12
94	account-v	247	25.02
95	determine-v	246	24.91
96	tend-v	244	24.71
97	want-v	244	24.71
98	move-v	242	24.51
99	interpret-v	237	24.00
100	explore-v	235	23.80

Appendix 3b: list of the top 100 verbs in the AAC

	Verb	Transcription (Meaning)	Raw Freq	Norm Freq
1	كان	/kaːna/ (be)	8376	848.84
2	قال	/qaːla/(say)	6852	694.39
3	تعالى	/taʕaːla/(exalt)	1689	171.17
4	جاء	/dʒaːʔa/(come/ to be mentioned)	1207	122.32
5	ورد	/warada/(has been mentioned/come)	1143	115.83
6	ذكر	/ðakara/(mention)	1005	101.85
7	وجد	/wadʒada/ (find/ exist)	934	94.65
8	رأى	/raʔa/ (see/ think/ perceive)	878	88.98
9	جعل	/dʒaʕala/(make/ create)	824	83.51
10	اراد	/ʔaraːda/(want)	694	70.33
11	قام	/qaːma/(base/ do)	606	61.41
12	وقع	/waqaʕa/(fall/ take place)	602	61.01
13	دل	/daˈla/(indicate)	594	60.20
14	يعلم	/jaʕlamu/(learn/ teach)	581	58.88
15	عرف	/ʕarafa/(know/define)	509	51.58
16	قدم	/qadama/(precede/offer)	508	51.48
17	أتى	/ʔata/ (come)	505	51.18
18	دخل	/daχala/(enter)	474	48.04
19	اشار	/ʔaʃaːra/ (point to/ mention)	460	46.62
20	اختلف	/ʔiχtalafa/(differ)	418	42.36
21	يعني	/jaʕni/(mean/ care/ pay attention)	404	40.94
22	ذهب	/ðahaba/(go)	404	40.94
23	وصف	/wasˤafa/(describe)	382	38.71
24	يعد	/juʕaˈdu/(consider/ is no longer/ count)	379	38.41
25	يمثل	/jumaˈθilu/ (represent)	371	37.60
26	استعمل	/ʔistaʕmala/(use)	352	35.67
27	اصبح	/ʔasˤbaħa/(become)	342	34.66
28	تم	/taˈma/(be completed)	326	33.04

	Verb	Transcription (Meaning)	Raw Freq	Norm Freq
51	كتب	/kataba/ (write)	213	21.59
52	يعمل	/jaʕmalu/(work/make/do)	213	21.59
53	اعتمد	/ʔiʕtamada/(depend/ rely on)	209	21.18
54	سمع	/samiʕa/ (hear)	202	20.47
55	سلم	/saˈlama/ (greet/ salute)	200	20.27
56	صلى	/sˤaˈla/ (pray)	194	19.66
57	تناول	/tanaːwala/ (deal with)	193	19.56
58	حاول	/ħaːwala/(try)	183	18.55
59	جرى	/dʒara/(run/ perform)	178	18.04
60	وصل	/wasˤala/ (arrive)	177	17.94
61	تبع	/tabaʕa/ (follow)	177	17.94
62	حدث	/ħadaθa/ (happen)	175	17.73
63	بقي	/baqja/ (stay/ remain)	173	17.53
64	بلغ	/balaɣa/ (reach)	170	17.23
65	يشمل	/jaʃmalu/ (include)	158	16.01
66	تحدث	/taħaˈdaθa/ (chat/ speak)	156	15.81
67	صح	/sˤaħa/(correct)	154	15.61
68	نزل	/nazala/(come down)	153	15.51
69	قصد	/qasˤada/ (intend/ aim)	152	15.40
70	يعود	/jaʕuːdu/(return)	152	15.40
71	اضاف	/ʔadˤaːfa/ (add)	150	15.20
72	يعبر	/juʕaˈbiru/(express)	144	14.59
73	زاد	/zaːda/(increase)	144	14.59
74	يقتضي	/jaqtadˤiː/(require)	143	14.49
75	جل	/dʒaˈla/ (be great)	143	14.49
76	يحتاج	/jaħtaːdʒu/(need)	140	14.19
77	يؤكد	/juːʔakidu/ (assure)	140	14.19
78	يشكل	/juʃaˈkilu/ (form)	140	14.19

	Verb	Transcription (Meaning)	Raw Freq	Norm Freq
29	تعلق	/taʕaˈlaqa/(be attached to)	309	31.31
30	خرج	/χaradʒa/ (go out)	307	31.11
31	ظهر	/ðˤahara/(appear/ to make sth appear)	304	30.81
32	يؤدي	/juʔaˈdiː/ (lead to/performs)	286	28.98
33	سمى	/saˈma/(call)	285	28.88
34	قرا	/qaraʔa/ (read)	281	28.48
35	حمل	/ħamala/(carry)	280	28.38
36	فعل	/faʕala/ (do)	276	27.97
37	جمع	/dʒamaʕa/ (combine/ agree on/ collect)	273	27.67
38	ثبت	/θabata/ (prove/ stay/to be fixed)	272	27.56
39	اخذ	/ʔaχaða/(take)	271	27.46
40	يبدو	/jabdu/(appear/seem)	260	26.35
41	بدأ	/badaʔa/(begin/ start (appear))	255	25.84
42	سبق	/sabaqa/(precede)	251	25.44
43	صار	/sˤaːra/(become)	249	25.23
44	يرتبط	/jartabitˤu/ (connect)	228	23.11
45	يلي	/jali/ (follow)	238	24.12
46	حذف	/ħaðafa/(omit)	237	24.02
47	يبين	/jubajinu/ (clarify)	222	22.50
48	وضع	/wadˤaʕa/ (put)	221	22.40
49	يفيد	/jufiːdu/(benefit)	220	22.30
50	استخدم	/ʔistaχdama/ (use)	219	22.19

	Verb	Transcription (Meaning)	Raw Freq	Norm Freq
79	يساعد	/jusaʔidu/(help)	135	13.68
80	زال	/zaːla/ (vanish/remain)	134	13.58
81	يعتبر	/juʕtabaru/ (consider)	131	13.28
82	نقل	/naqala/ (report)	130	13.17
83	ينظر	/janðˤuru/ (see/ look)	119	12.06
84	يعيش	/jaʕiːʃu/ (live)	118	11.96
85	دعا	/daʕa/ (call)	114	11.55
86	يدرس	/jadrusu/ (study)	112	11.35
87	يتطلب	/jatatˤaˈlabu/ (demand)	108	10.94
88	يتضمن	/jatadˤaˈmanu/ (include)	107	10.84
89	ينطق	/jantˤiqu/ (utter)	106	10.74
90	يفهم	/jafhamu/ (understand)	105	10.64
91	عرض	/ʕaradˤa/ (show)	105	10.64
92	رحم	/raħima/(have mercy on)	102	10.34
93	يميز	/juˈmajizu/(distinguish)	101	10.24
94	حصل	/ħasˤala/ (happen/occur)	100	10.13
95	عز	/ʕaˈza/ (glorify)	97	9.83
96	يقابل	/juqaːbilu/ (confront)	94	9.53
97	يضم	/jadˤuˈmu/ (include)	94	9.53
98	سار	/sara/ (walk)	93	9.42
99	اتفق	/ʔiˈtafaqa/ (agree)	93	9.42
100	خالف	/χaːlafa/ (disagree)	92	9.32

Appendix 4: List of verbs in the selected clause structures in expert writers' corpora

1. List of verbs in the AEC

Copular

Copular	Raw	Norm	Faithfulness
Become	582	59	98
seem	544	55	97
be	463	47	46
appear	324	33	55
remain	249	25	81
get	93	9	30
turn	55	6	20
look	43	4	12
go	36	4	7
come	14	1	2
make	12	1	1
Total	2415	245	Mean=41

Transitive

Transitive	Raw	Norm	Faithfulness
include	975	99	98
show	941	95	94
provide	900	91	90
use	892	90	89
see	869	88	87
involve	816	83	100
find	793	80	81
suggest	656	66	98
discuss	601	61	98
represent	557	56	90
make	519	53	52
express	505	51	97
argue	497	50	79
indicate	472	48	92
describe	468	47	85
identify	454	46	81
need	454	46	98
examine	454	46	100
follow	436	44	44

Transitive	Raw	Norm	Faithfulness
take	428	43	43
consider	425	43	65
say	425	43	77
produce	421	43	98
illustrate	416	42	98
present	406	41	80
note	404	41	96
construct	390	39	80
create	385	39	91
explain	381	39	95
require	357	36	83
contain	355	36	99
reflect	351	36	90
propose	328	33	82
demonstrate	326	33	99
understand	325	33	75
mean	322	33	54
establish	306	31	92
give	298	30	30
report	298	30	72
reveal	291	29	95
observe	288	29	97
support	285	29	96
assume	282	29	81
address	277	28	94
know	271	27	54
offer	269	27	85
develop	264	27	70
have	260	26	26
analyse	256	26	88
tend	244	25	100
claim	236	24	93
explore	235	24	100
define	232	23	66
determine	232	23	94
want	217	22	89
do	209	21	21
introduce	206	21	81
share	191	19	71
mark	172	17	68
allow	168	17	36
expect	161	16	54
write	159	16	38
help	158	16	58

Transitive	Raw	Norm	Faithfulness
speak	154	16	21
begin	153	15	60
compare	139	14	41
interpret	133	13	56
learn	129	13	36
ask	129	13	42
get	126	13	41
think	117	12	39
come	108	11	16
play	105	11	37
start	96	10	38
apply	93	9	22
increase	75	8	30
draw	70	7	17
set	47	5	17
contribute	44	4	16
relate	28	3	7
move	26	3	11
lead	22	2	5
regard	11	1	3
go	9	1	2
turn	8	1	3
call	7	1	2
Total	26998	2734	Mean=65

Phrasal type 2

Phrasal type 2	Raw	Norm	Faithfulness
take up	56	6	6
take on	32	3	3
set up	31	3	12
make up	21	2	2
find out	17	2	2
set Out	13	1	5
work out	12	1	4
take Over	9	1	1
start Off	5	1	2
follow up	4	0	0
start out	4	0	2
get in	3	0	1
get out	3	0	1
turn off	3	0	1
turn up	3	0	1

Phrasal type 2	Raw	Norm	Faithfulness
draw up	2	0	0
move away	2	0	1
set apart	2	0	1
set aside	2	0	1
set off	2	0	1
show off	2	0	0
take away	2	0	0
take forward	2	0	0
take in	2	0	0
turn back	2	0	1
call out	1	0	0
get cross	1	0	0
get through	1	0	1
give away	1	0	0
give in	1	0	0
give up	1	0	0
lead on	1	0	0
mark out	1	0	0
set forth	1	0	0
take aback	1	0	0
think over	1	0	0
turn down	1	0	0
turn on	1	0	0
turn round	1	0	0
write down	1	0	0
write out	1	0	0
Total	252	26	Mean=1

Prepositional type 1

Prepositional type 1	Raw	Norm	Faithfulness
refer to	572	58	96
focus on	462	47	79
lead to	283	29	65
account for	247	25	100
draw on	243	25	57
contribute to	217	22	81
look at	202	20	56
relate to	127	13	31
turn to	103	10	38
apply to	91	9	21
allow for	88	9	19

Prepositional type 1	Raw	Norm	Faithfulness
argue for	69	7	11
start with	49	5	19
begin with	48	5	19
call for	48	5	16
think about	39	4	13
reflect on	38	4	10
work on	37	4	11
come to	33	3	5
speak to	30	3	4
look for	26	3	7
report on	25	3	6
come into	24	2	4
think of	24	2	8
speak in	20	2	3
speak of	19	2	3
write about	19	2	5
come with	18	2	3
ask for	17	2	6
develop from	17	2	5
start from	16	2	6
learn about	14	1	4
argue against	12	1	2
learn from	12	1	3
begin by	11	1	4
identify with	11	1	2
develop into	9	1	2
work for	9	1	3
apply for	8	1	2
play with	8	1	3
speak about	8	1	1
apply in	7	1	2
know about	7	1	1
go through	17	2	3
go for	6	1	1
go into	6	1	1
provide for	6	1	1
help with	5	1	2
learn of	5	1	1
ask about	4	0	1
come in	4	0	1
go with	4	0	1
play on	4	0	1
speak with	4	0	1
start by	4	0	2

Prepositional type 1	Raw	Norm	Faithfulness
develop as	3	0	1
get into	3	0	1
get on	3	0	1
go against	3	0	1
help in	3	0	1
speak on	3	0	0
write on	3	0	1
accord with	2	0	1
come across	2	0	0
come by	2	0	0
get round	2	0	1
go over	2	0	0
speak for	2	0	0
argue with	1	0	0
come under	1	0	0
get at	1	0	0
get to	1	0	0
go about	1	0	0
go to	1	0	0
speak against	1	0	0
Total	3476	352	Mean=11

Phrasal prepositional type 1

Phrasal prepositional type 1	Raw	Norm	Faithfulness
come up with	10	1	1
move away from	7	1	1
come back to	5	1	0
go along with	5	1	1
follow up on	4	0	1
look forward to	4	0	0
refer back to	4	0	1
develop out of	3	0	1
get on with	3	0	1
go out of	3	0	1
lead up to	3	0	1
Look back at	3	0	1
Look up at	3	0	1
start off with	3	0	1
start out with	3	0	1
write back to	3	0	1
come out with	2	0	1

Phrasal prepositional type 1	Raw	Norm	Faithfulness
come up against	2	0	0
get down with	2	0	0
get away with	1	0	1
get out of	1	0	1
get up to	1	0	0
go up to	1	0	0
move along with	1	0	0
begin out of	1	0	3
Total	78	8	Mean=1

Complex copular

Complex copular	Raw	Norm	Faithfulness
make	334	34	33
call	229	23	75
consider	227	23	35
allow	210	21	44
take	139	14	14
find	139	14	14
see	127	13	13
regard	105	11	33
interpret	103	10	43
expect	101	10	34
help	100	10	36
define	94	10	27
understand	89	9	21
identify	85	9	15
construct	83	8	17
lead	76	8	17
describe	75	8	14
know	74	7	15
use	71	7	7
present	69	7	14
require	65	7	15
assume	61	6	18
represent	56	6	9
say	51	5	9
mean	49	5	8
think	49	5	16
show	37	4	4
analyse	29	3	10
report	25	3	6

Complex copular	Raw	Norm	Faithfulness
mark	24	2	9
include	19	2	2
establish	15	2	5
discuss	11	1	2
propose	11	1	3
explain	9	1	2
get	9	1	3
offer	7	1	2
introduce	6	1	2
address	5	1	2
reveal	4	0	1
create	3	0	1
suggest	2	0	0
observe	2	0	1
demonstrate	1	0	0
Total	3080	312	Mean=15

Complex transitive

Complex transitive	Raw	Norm	Faithfulness
take	77	8	8
draw	56	6	13
make	4	0	0
get	1	0	0
Total	138	14	Mean=5

Ditransitive/Double object

Ditransitive/ Double object	Raw	Norm	Faithfulness
give	137	14	14
ask	114	12	37
offer	19	2	6
show	13	1	1
get	6	1	2
take	5	1	1
allow	2	0	0
write	2	0	0
draw	1	0	0
Total	299	30	Mean=7

Prepositional type 2a

Prepositional type 2a	Raw	Norm	Faithfulness
give to	83	8	8
offer to	10	1	3
write to	7	1	2
offer for	7	1	2
ask to	7	1	2
show to	2	0	0
Total	116	12	Mean=3

Prepositional type 2b

Prepositional type 2b	Raw	Norm	Faithfulness
create for	26	3	6
provide for	18	2	2
set to	17	2	6
Provide to	16	2	2
say to	13	1	2
present for	12	1	3
address to	11	1	4
know to	10	1	2
mean for	8	1	2
seem to	8	1	1
produce for	6	1	1
explain to	6	1	2
find for	4	0	0
appear to	4	0	1
report to	3	0	1
reveal to	3	0	1
suggest to	2	0	0
discuss with	2	0	0
mean to	2	0	0
present to	2	0	0
demonstrate to	2	0	1
represent to	1	0	0
explain for	1	0	0
Total	177	18	Mean=2

Prepositional type 3

Prepositional type 3	Raw	Norm	Faithfulness
play a role in	134	14	47

Prepositional type 3	Raw	Norm	Faithfulness
give rise to	80	8	8
take sth into account	62	6	6
make use of	44	4	4
draw attention to	43	4	10
make reference to	13	1	1
take part in	13	1	1
focus our attention on	12	1	2
make sense of	11	1	1
call sth into question	11	1	4
give way to	10	1	1
take advantage of	10	1	1
take sth into consideration	10	1	1
play a part in	9	1	3
take account of	8	1	1
turn attention to	8	1	3
take responsibility for	7	1	1
come to grips with	5	1	1
take care of	4	0	0
take stock of	4	0	0
get rid of	3	0	1
make fun of	3	0	0
call to mind	3	0	1
call attention to	2	0	1
take control of	2	0	0
take hold of	2	0	0
take precedence over	2	0	0
come to terms with	1	0	0
get hold of	1	0	0
Total	517	52	Mean=3

Prepositional type 4a

Prepositional type 4a	Raw	Norm	Faithfulness
provide someone with	47	5	5
introduce someone to something	16	2	6
present someone with	16	2	3
lead someone to	14	1	3
ask someone about	11	1	4
ask someone for	8	1	3
focus someone on	6	1	1
refer someone to sth	6	1	1
Total	124	13	Mean=3

Prepositional 4b

Prepositional type 4b	Raw	Norm	Faithfulness
base on	483	49	91
associate sth with sth	366	37	92
relate sth to sth	230	23	56
compare sth to sth	109	11	32
compare sth with sth	86	9	26
apply sth to sth	67	7	16
know sth about	36	4	7
introduce sth to sth	24	2	9
turn sth into	23	2	8
say sth about	19	2	3
expect sth from sth	10	1	3
make sth of sth	8	1	1
reveal sth about sth	7	1	2
contribute sth to sth	7	1	3
make sth into sth	5	1	1
start sth with sth	3	0	1
make sth to sth	2	0	0
allow space/time for	2	0	0
make sth from sth	1	0	0
Total	1488	151	Mean=18

Phrasal prepositional type 2

Phrasal prepositional type 2	Raw	Norm	Faithfulness
made up of	16	2	2
turn it away from	2	0	1
follow it up with	2	0	0
lead the discourse away from	1	0	0
set it apart from the views	1	0	0
take it out of its context	1	0	0
Total	23	2	Mean=1

Phrasal prepositional type 3

Phrasal prepositional type 3	Raw	Norm	Faithfulness
set them/sth apart form	4	0	2
get sth/someone back on	2	0	1
take sth out of	2	0	0
take us away from	1	0	0

Phrasal prepositional type 3	Raw	Norm	Faithfulness
give herself up to the ride	1	0	0
Total	10	1	Mean=1

2. List of verbs in the AAC

Copular

Copular	Raw	Norm	faithfulness
/kana/ (be)	805	82	81
/ʔsˤbaha/(become)	311	32	91
/sˤa:ra/(become)	238	24	96
/yabdu/(appear/seem)	118	12	45
/ʔata/ (come)	89	9	18
/baqia/ (stay/ remain)	88	9	51
/zala/ (vanish/remain)	71	7	53
/yaʕud/ (consider/ is no longer/ count)	58	6	15
/ðˤahara/ (appear/ to make sth appear)	8	1	3
Total	1786	181	Mean=50

Transitive

Transitive	Raw	Norm	faithfulness
/ðakara/(mention)	995	101	99
/qala/(say)	955	97	96
/ʔarada/(want)	694	70	100
/wadʒada/ (find/ exist)	566	57	61
/yaʕni:/ (mean/ care/ pay attention)	390	40	97
/istaʕmala/ (use)	352	36	100
/yaʕlamu/ (learn/ teach)	334	34	62
/ʕarafa/ (know/define)	318	32	62
/yumaθılu/(represent)	286	29	77
/qarʔa/ (read)	281	28	100
/ħamala/ (carry)	280	28	100
/raʔa/(see/ think/ perceive)	246	25	28
/ħaðafat/(omit)	237	24	100
/qadama/ (precede/offer)	221	22	44
/yufi:du/(benefit)	220	22	100
/istaxdama/ (use)	219	22	100
/tanawala/ (deal with)	193	20	100
/ʔaxaða/(take)	191	19	70
/samiʕa/ (hear)	191	19	95

Transitive	Raw	Norm	faithfulness
/ħawala/(try)	183	19	100
/wadˁaʕa/ (put)	174	18	79
/tabaʕa/ (follow)	165	17	93
/yubayınu/(clarify/ explain)	164	17	74
/kataba/ (write)	162	16	76
/dʒamaʕa/ (combine/ agree on/ collect)	155	16	57
/θabata/ (prove/ stay/to be fixed)	152	15	56
/warada/(has been mentioned/come)	151	15	15
/qasˁada/ (intend/ aim)	151	15	99
/yaqtadˁi:/(require)	143	14	100
/yuʔdi:/ (lead to/performs)	141	14	49
/yuʃakılu/ (form)	136	14	97
/balaɣa/ (reach)	121	12	71
/naqala/ (report)	121	12	93
/faʕala/ (do)	118	12	43
/wasˁafa/ (describe)	117	12	31
/yu:ʔkıdu/ (assure)	116	12	83
/yadrusu/ (study)	112	11	100
/yaʃmalu/ (include)	108	11	68
/yatatˁalabu/(demand)	108	11	100
/yatadˁamnu/ (include)	107	11	100
/yantˁiqu/ (utter)	106	11	100
/yafhamu/ (understand)	105	11	100
/raħəma/(have mercy on)	102	10	100
/ʕaradˁa/ (show)	100	10	95
/yuqabilu/ (confront)	94	10	100
/nazala/ (come down)	92	9	64
/χalafa/ (disagree)	89	9	97
/yadˁumu/ (include)	76	8	81
/iʕtamada/ (depend/ rely on)	74	7	35
/badaʔ/(begin/ start (appear))	71	7	28
/yumaizu/ (distinguish)	69	7	68
/dʒaʔa/(come/ be mentioned)	63	6	6
/xaradʒa/ (go out)	50	5	16
/zada/(increase)	43	4	30
/yalı/ (follow)	39	4	16
/ðˁahara/ (appear/ to make sth appear)	38	4	13
/yanðˁuru/ (see/ look)	36	4	30
/ʔdˁafa/ (add)	34	3	23
/wasˁala/ (arrive)	32	3	18
/yaʕmalu/ (work/make/do)	27	3	13
/sabaqa/(precede)	21	2	8
/daʕa/ (call)	21	2	18
/yartabıtˁu/ (connect)	18	2	8

Transitive	Raw	Norm	faithfulness
/dʒara/(run/ perform)	18	2	10
/ʔʃara/ (point to/ mention)	17	2	4
/yusæʔɪdu/(help)	13	1	10
/ʔata/ (come)	12	1	2
/yuʕadu/ (consider/ is no longer/ count)	12	1	3
/daxala/ (enter)	9	1	2
/yaħtadʒu/(need)	8	1	6
/ħadaθa/ (happen)	3	0	2
/taħadaθa/ (chat/ speak)	2	0	1
/yabdu/(appear/ seem)	1	0	0
Total	11569	1172	Mean=59

Prepositional type 1

Prepositional type 1	Raw	Norm	Faithfulness
/dalla/ (indicate) ʕala (on)	582	59	98
/ʔʃara/ (point to/ mention) ila (to)	443	45	96
/ðahaba/(go) ila (to)	310	31	77
/taʕlaqa/ (be attached to) bi (with)	262	27	85
/yuʔdi:/ (lead to) ila (to)	145	15	51
/iʕtamada/ (depend/ rely on) ʕala (on)	135	14	65
/daxala/ (enter) ʕala (on)	134	14	28
/yaħtadʒu/ (need) ila (to)	131	13	94
/yaʕu:du/(return) ila (to)	126	13	83
/yuʕabɪru/ (express) ʕan (about)	123	12	85
/wasʿala/ (arrive) ila (to)	119	12	67
/yartabɪtʿu/ (connect) bi (to)	117	12	51
/ʔdʿafa/ (add) ila (to)	108	11	72
/ixtalafa/ (differ) ʕan (from)	102	10	24
/waqaʕa/ (fall/ take place) fi (in)	98	10	16
/taħadaθa/ (chat/ speak) ʔan (about)	96	10	62
/dʒamaʕa/ (agree on) ʕala (on)	87	9	32
/dʒaʔa/ (come/ mentioned) bi (with)	80	8	8
/xaradʒa/ (go out) ʕan (from)	79	8	26
/yumaθɪlu/(represent) fi (in)	77	8	21
/yaʕmalu/ (work/make/do) ʕala (on)	76	8	36
/qama/ (do) bi	73	7	12
/yanðʿuru/ (see/ look) ila (to/at)	65	7	55
/waqaʕa/ (fall/ take place) ʕala (on)	64	6	11
/badaʔ/(begin/ start (appear)) bi (with)	63	6	25
/yusæʔɪdu/ (help) ʕala (on)	58	6	43
/yartabɪtʿu/ (connect) byna (between)	55	6	24

Prepositional type 1	Raw	Norm	Faithfulness
/daʕa/ (call) ila (to)	54	5	47
/yaʃmalu/ (include) ʔala (on)	50	5	32
/qama/ (base) ʕala(on)	44	4	7
/ʔata/ (come) bi (with)	39	4	8
/itafaqa/ (agree) maʕa (with)	34	3	37
/xaradʒa/ (go out) ala (of)	33	3	11
/itafaqa/(agree) ʕala (on)	32	3	34
/dʒamaʕa/ (combine) byna(between)	31	3	11
/yaʕu:du/(return) ʔala (on)	25	3	16
/ʔaxaða/(take) bi	25	3	9
/yu:ʔkıdu/ (assure)	24	2	17
/xaradʒa/ (go out) bi (with)	19	2	6
/yumaizu/ (distinguish) byna (between)	16	2	16
/zada/(increase) min (from)	13	1	9
/yuscæʔıdu/(help) fi (in)	12	1	9
/yanðʕuru/ (see/ look) fi (in)	12	1	10
/itafaqa/(agree) fi (in)	6	1	6
/zada/(increase) fi (in)	4	0	3
/itafaqa/(agree) ħawla	1	0	1
Total	4282	434	Mean=36

Complex copular

Complex copular	Raw	Norm	faithfulness
/dʒaʕala/(make/ create)	824	84	100
/raʔa/(see/ think/ perceive)	548	56	62
/wadʒada/ (find/ exist)	330	33	35
/yuʕadu/(consider/ is no longer/ count)	289	29	76
/samma/ (call)	285	29	100
/ʕarafa/ (know/define)	191	19	38
/yuʕtabaru/ (consider)	130	13	99
/wasʕafa/(describe)	56	6	15
/daʕa/ (call)	2	0	2
Total	2655	269	Mean=59

Complex transitive

Complex transitive	Raw	Norm	faithfulness
/wadʕaʕa/ (put)	40	4	18
/daxala/ (enter)	26	3	5
Total	66	7	Mean= 12

Ditransitive/ Double object

Ditransitive/ Double object	Raw	Norm	faithfulness
/zada/ (increase)	10	1	7
/yaʕlamu/ (learn/ teach)	3	0	1
Total	13	1	Mean=4

Prepositional type 2a

Prepositional type 2a	Raw	Norm	faithfulness
Total	0	0	Mean=0

Prepositional type 2b

Prepositional type 2b	Raw	Norm	faithfulness
/yubayınu/(clarify/ explain)	47	5	21
/yabdu/(appear/seem)	34	3	13
/qadama/ (precede/offer)	26	3	5
/ðˤahara/ (appear/ to make sth appear)	24	2	8
/kataba/ (write)	18	2	8
/qala/(say)	14	1	1
/ʕaradˤa/ (show)	5	1	5
/naqala/ (report)	4	0	3
Total	172	17	Mean=8

Prepositional type 3

Prepositional type 3	Raw	Norm	faithfulness
/ʔaxaða/(take) /bi/(into)/ʔliʕtibar/(consideration)	8	1	3
/ʔaxaða/(take) /bi/(into)/ʔlħusban/(account)	8	1	3
Total	16	2	Mean=3

Prepositional type 4a

Prepositional type 4a	Raw	Norm	faithfulness
/yusæʔıdu/(help) ʕala (on)	45	5	33

Prepositional type 4a	Raw	Norm	faithfulness
/daʕa/ (call) ila (to)	24	2	21
/yartabɪtˁu/ (connect) bi (with)	2	0	1
/yadˁumu / (include) ila (to)	2	0	2
Total	73	7	Mean= 14

Prepositional type 4b

Prepositional type 4b	Raw	Norm	faithfulness
/yartabɪtˁu/ (connect) bi (with)	35	4	15
/yusæʔɪdu/(help) ʕala (on)	6	1	4
/yadˁumu / (include) ila (to)	11	1	12
/daʕa/ (call) ila (to)	1	0	1
Total	53	5	Mean=8

Appendix 5a: Ethical Approval for the NNC

University of Reading

Memo

School of Literature and Languages
Department of English Language and Applied Linguistics

Please reply to: Dr. Clare Wright (c.e.m.wright@reading.ac.uk)

To Dr. Sylvia Jaworska

From Dr. Clare Wright

Copy Manal Alangari
Anne Whitbread (file)

Date 18th May 2016

Your application for Ethical Approval

Your project entitled **'A corpus-based study of collocations in the writing of advanced Saudi learners of English'** has been considered by the School Ethics Committee, and we are pleased to report that the Committee has raised no ethical objections and subject to your undertaking to store the consent forms in the Department Office in the normal way, it has accordingly given permission for the project to proceed under the exceptions procedure as outlined in paragraph 6 of the *University's Ethics Guidance to Schools.*

Signed

Dr. Clare Wright

On behalf of the School Ethics Committee,
Prof. Catherine Leglu, School Director of Research.
Prof. Alison Donnell, Head of School.

Appendix 5b: Ethical Approval for the NSC

University of Reading

Memo

School of Literature and Languages
Department of English Language and Applied Linguistics

Please reply to: Dr. Clare Wright (c.e.m.wright@reading.ac.uk)

To Dr. Sylvia Jaworska

From Dr. Clare Wright

Copy Manal Alangari
Anne Whitbread (file)

Date 11th February 2016

Your application for Ethical Approval

Your project entitled **'A corpus-based study of collocations in the writing of Advanced Saudi Learners of English'** has been considered by the School Ethics Committee, and we are pleased to report that the Committee has raised no ethical objections and subject to your undertaking to store the consent forms in the Department Office in the normal way, it has accordingly given permission for the project to proceed under the exceptions procedure as outlined in paragraph 6 of the *University's Ethics Guidance to Schools.*

Signed

Dr. Clare Wright

On behalf of the School Ethics Committee,
Prof. Catherine Leglu, School Director of Research.
Prof. Alison Donnell, Head of School.

Appendix 5c: information sheet

University of Reading

Department of English Language and Applied Linguistics

HumSS Building
The University of Reading
Whiteknights, PO Box 218
Reading RG6 6AA

Phone 01183788141
+44 (0)118 378 6472
+44 (0)118 975 6506
Email
*appling@reading.ac.uk*p.a.thompson@reading.ac.uk

Researcher:
Manal Alangari
Phone:
UK:07467523872
SA:0506853224
Email:
M.A.M.Alangari@pgr.reading.ac.uk

Supervisor:
Dr.Sylvia Jaworska
Phone: 01183787885
Email: s.jaworska@reading.ac.uk

INFORMATION SHEET

The purpose of this research is to investigate the use of verb-noun collocations in the academic writing of Saudi learners of English. The researcher is interested in identifying the syntactic and the semantic collocational patterns most commonly used by Saudi learners and discovering the influence of the learners' first language (Arabic) on their production.

Participants are drawn from students at the English Language and Literature department at the University of Dammam. Specifically, tests done for applied linguistics modules will be anonymized, typed, and uploaded to computer software for further analysis. The data will be collected and stored by the researcher herself on a password-protected computer and software. All confidential information (e.g. students names' and contact details) will be available to the researcher only and will be kept securely by the researcher only to be used to make further contact with the participants in the second phase of the study. All data and information are to be used for research purposes only. Participants may withdraw at any time or any stage of the study by contacting the researcher at the above given contact details.

This project has been subject to ethical review by the School Ethics and Research Committee, and has been allowed to proceed under the exceptions procedure as outlined in paragraph 6 of the University's *Notes for Guidance* on research ethics.

If you have any queries or wish to clarify anything about the study, please feel free to contact my supervisor at the address above or by email at [include supervisor's email address here]

Signed

Manal Alangari

PhD student

Appendix 5d: consent forms for the NNC

School of Literature and Languages
Department of English Language and Applied Linguistics

Coursework permission form

I consent to any written work I produce as part of my programme of study in the Department of English Language and Applied Linguistics in the School of Literature and Languages, the University of Reading, being used anonymously for academic literacy advice input for other students and/or research into academic writing.

Date:

Name:

Programme:

Signature:

Appendix 5e: consent forms for the NSC

School of Literature and Languages
Department of English Language and Applied Linguistics

University of Reading

ETHICS COMMITTEE

Consent Form

Project title: A corpus-based study of collocations in the writing of Advanced Saudi Learners of English

I understand the purpose of this research and understand what is required of me; I have read and understood the Information Sheet relating to this project, which has been explained to me by Manal Alangari. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project at any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix 6a: the 100 most frequent verbs in the NNC

	Lempos	Raw Freq	Norm Freq
1	be	2408	5461.56
2	have	360	816.51
3	use	260	589.70
4	do	157	356.09
5	find	91	206.40
6	see	75	170.11
7	define	65	147.43
8	make	64	145.16
9	suggest	60	136.09
10	speak	57	129.28
11	show	55	124.74
12	include	54	122.48
13	change	53	120.21
14	learn	52	117.94
15	create	51	115.67
16	consider	50	113.40
17	develop	48	108.87
18	write	47	106.60
19	describe	45	102.06
20	take	45	102.06
21	give	45	102.06
22	mean	43	97.53
23	base	43	97.53
24	involve	43	97.53
25	become	42	95.26
26	argue	41	92.99
27	know	40	90.72
28	allow	39	88.46
29	believe	39	88.46
30	add	37	83.92
31	discuss	36	81.65
32	read	36	81.65
33	explain	36	81.65
34	begin	34	77.11
35	reflect	34	77.11
36	demonstrate	33	74.85
37	study	32	72.58
38	look	32	72.58
39	go	29	65.77
40	occur	29	65.77
41	require	28	63.51
42	form	28	63.51

	Lempos	Raw Freq	Norm Freq
51	highlight	24	54.43
52	focus	24	54.43
53	apply	24	54.43
54	represent	24	54.43
55	compare	24	54.43
56	claim	24	54.43
57	combine	24	54.43
58	identify	23	52.17
59	note	23	52.17
60	follow	23	52.17
61	tend	21	47.63
62	accord	21	47.63
63	feel	20	45.36
64	conclude	20	45.36
65	express	20	45.36
66	produce	19	43.09
67	propose	19	43.09
68	relate	19	43.09
69	establish	18	40.83
70	provide	18	40.83
71	view	18	40.83
72	differ	18	40.83
73	contain	17	38.56
74	encode	17	38.56
75	explore	17	38.56
76	attach	17	38.56
77	come	17	38.56
78	walk	17	38.56
79	measure	17	38.56
80	influence	17	38.56
81	attend	16	36.29
82	ask	16	36.29
83	alter	16	36.29
84	work	16	36.29
85	lead	16	36.29
86	hold	16	36.29
87	complete	16	36.29
88	examine	15	34.02
89	support	14	31.75
90	think	14	31.75
91	exist	13	29.49
92	get	12	27.22

	Lempos	Raw Freq	Norm Freq
43	need	27	61.24
44	depend	27	61.24
45	affect	27	61.24
46	refer	26	58.97
47	teach	25	56.70
48	want	25	56.70
49	say	25	56.70
50	understand	25	56.70

	Lempos	Raw Freq	Norm Freq
93	happen	12	27.22
94	help	11	24.95
95	seem	11	24.95
96	acquire	11	24.95
97	appear	10	22.68
98	marry	10	22.68
99	choose	10	22.68
100	start	9	20.41

Appendix 6b: the 100 most frequent verbs in the NSC

	Lempos	Raw Freq	Norm Freq
1	be	1,917	4676.86
2	use	410	1000.27
3	have	341	831.93
4	learn	223	544.05
5	do	172	419.62
6	speak	171	417.19
7	say	133	324.48
8	read	100	243.97
9	mean	99	241.53
10	refer	73	178.10
11	talk	69	168.34
12	show	67	163.46
13	make	62	151.26
14	know	61	148.82
15	understand	59	143.94
16	pronounce	55	134.18
17	acquire	52	126.86
18	hear	50	121.98
19	distinguish	49	119.54
20	give	49	119.54
21	process	44	107.35
22	accord	43	104.91
23	parse	39	95.15
24	ask	38	92.71
25	reach	36	87.83
26	go	36	87.83
27	apply	36	87.83
28	code	35	85.39
29	call	35	85.39
30	seem	33	80.51
31	come	33	80.51
32	depend	33	80.51
33	tend	32	78.07
34	write	32	78.07
35	see	31	75.63
36	teach	31	75.63
37	discover	30	73.19
38	differ	30	73.19
39	help	30	73.19
40	produce	29	70.75
41	want	29	70.75
42	relate	28	68.31
43	drop	27	65.87
44	choose	26	63.43

	Lempos	Raw Freq	Norm Freq
51	take	24	58.55
52	name	23	56.11
53	occur	23	56.11
54	include	23	56.11
55	focus	23	56.11
56	marry	22	53.67
57	identify	22	53.67
58	get	22	53.67
59	listen	22	53.67
60	tell	21	51.23
61	solve	21	51.23
62	exist	21	51.23
63	like	21	51.23
64	try	20	48.79
65	educate	18	43.91
66	base	17	41.47
67	consider	17	41.47
68	work	17	41.47
69	reflect	16	39.03
70	chain	16	39.03
71	answer	16	39.03
72	start	16	39.03
73	determine	15	36.60
74	suggest	15	36.60
75	repeat	15	36.60
76	form	15	36.60
77	appear	15	36.60
78	support	14	34.16
79	operate	14	34.16
80	switch	14	34.16
81	study	14	34.16
82	become	14	34.16
83	belong	13	31.72
84	activate	13	31.72
85	shadow	13	31.72
86	sound	13	31.72
87	connect	13	31.72
88	need	13	31.72
89	practice	13	31.72
90	denote	12	29.28
91	turn	12	29.28
92	organize	12	29.28
93	begin	12	29.28
94	believe	12	29.28

	Lempos	Raw Freq	Norm Freq
45	bear	25	60.99
46	claim	25	60.99
47	think	24	58.55
48	happen	24	58.55
49	change	24	58.55
50	find	24	58.55

	Lempos	Raw Freq	Norm Freq
95	generate	12	29.28
96	indicate	11	26.84
97	affect	11	26.84
98	attract	11	26.84
99	spell	11	26.84
100	label	11	26.84

Appendix 7: List of verbs in the selected clause structures in novice writers' corpora

1. List of verbs in the NNC

Copular

Copular	Raw	Norm	Faithfulness
be	551	1250	55
become	42	95	100
feel	18	41	90
seem	11	25	100
appear	7	16	70
get	2	5	17
Total	631	1431	Mean= 72

Transitive

Transitive	Raw	Norm	Faithfulness
use	237	538	91
have	190	431	53
find	75	170	82
suggest	59	134	98
include	54	122	100
show	52	118	95
create	51	116	100
see	50	113	67
define	45	102	69
develop	42	95	88
argue	40	91	98
learn	39	88	75
mean	39	88	91
describe	38	86	84
discuss	36	82	100
explain	36	82	100
involve	35	79	81
demonstrate	33	75	100
make	33	75	52
believe	29	66	74
change	29	66	55

2. List of verbs in the NSC

Copular

Copular	Raw	Norm	Faithfulness
be	649	1583	65
seem	24	59	73
become	10	24	71
get	5	12	23
appear	5	12	33
sound	3	7	23
come	1	2	3
turn	1	2	8
Total	698	1703	Mean= 37

Transitive

Transitive	Raw	Norm	Faithfulness
use	393	959	96
have	276	673	81
say	122	298	92
learn	106	259	48
mean	72	176	73
read	68	166	68
pronounce	52	127	95
hear	50	122	100
acquire	50	122	96
understand	47	115	80
show	46	112	69
know	45	110	74
process	42	102	95
speak	34	83	20
distinguish	31	76	63
discover	30	73	100
make	30	73	48
reach	28	68	78
produce	27	66	93
teach	27	66	87
choose	26	63	100

Transitive	Raw	Norm	Faithfulness
study	29	66	91
require	28	64	100
affect	27	61	100
need	27	61	100
form	25	57	89
want	25	57	100
do	25	57	16
reflect	24	54	71
represent	24	54	100
read	24	54	67
understand	23	52	92
consider	22	50	44
highlight	22	50	92
note	22	50	96
begin	21	48	62
know	21	48	53
speak	21	48	37
tend	21	48	100
claim	20	45	83
express	19	43	95
produce	19	43	100
propose	19	43	100
say	19	43	76
contain	17	39	100
explore	17	39	100
measure	17	39	100
teach	17	39	68
write	17	39	36
attend	16	36	100
hold	16	36	100
take	16	36	36
alter	16	36	100
encode	15	34	88
examine	15	34	100
identify	15	34	65
influence	15	34	88
add	14	32	38
follow	13	29	57
give	13	29	29
establish	11	25	61
choose	10	23	100
marry	10	23	100
support	10	23	71
acquire	9	20	82
complete	8	18	50
help	8	18	73

Transitive	Raw	Norm	Faithfulness
do	26	63	15
see	26	63	84
want	26	63	90
claim	25	61	100
tend	25	61	78
include	23	56	100
find	22	54	92
drop	20	49	74
identify	20	49	91
give	19	46	39
try	18	44	90
change	16	39	67
reflect	16	39	100
help	15	37	50
determine	14	34	93
suggest	14	34	93
support	14	34	100
get	13	32	59
activate	12	29	92
need	12	29	92
repeat	12	29	80
affect	11	27	100
answer	11	27	69
attract	11	27	100
believe	11	27	92
form	11	27	73
generate	11	27	92
indicate	11	27	100
solve	10	24	48
study	10	24	71
take	9	22	38
apply	8	20	22
ask	8	20	21
like	8	20	38
organize	7	17	58
practice	7	17	54
think	7	17	29
connect	6	15	46
educate	6	15	33
marry	6	15	27
start	6	15	38
write	6	15	19
denote	5	12	42
label	4	10	36
parse	4	10	10
spell	4	10	36

Transitive	Raw	Norm	Faithfulness
lead	8	18	50
provide	8	18	44
compare	7	16	29
get	7	16	58
start	7	16	78
allow	6	14	15
ask	6	14	38
view	6	14	33
apply	4	9	17
combine	4	9	17
attach	3	7	18
conclude	3	7	15
think	3	7	21
come	1	2	6
look	1	2	3
total	2158	4895	Mean= 71

Transitive	Raw	Norm	Faithfulness
call	2	5	6
name	2	5	9
switch	2	5	14
talk	2	5	3
tell	2	5	10
go	1	2	3
Total	2199	5365	Mean= 64

Phrasal type 2

Phrasal type 2	Raw	Norm	Faithfulness
find out	2	5	2
get across	1	2	8
make up	1	2	2
take over	1	2	2
write down	1	2	2
Total	6	14	Mean= 3

Phrasal type 2	Raw	Norm	Faithfulness
find out	1	2	4
write down	1	2	3
Total	2	5	Mean= 4

Prepositional type 1

Prepositional type 1	Raw	Norm	Faithfulness
look at	27	61	84
depend on	26	59	96
refer to	26	59	100
focus on	22	50	92
relate to	17	39	89
combine with	11	25	46
attach to	9	20	53
come from	8	18	46
allow for	8	18	21
apply to	6	14	25
lead to	6	14	38

Prepositional type 1	Raw	Norm	Faithfulness
refer to	69	168	95
talk to	32	78	46
speak to	30	73	17
depend on	28	68	85
focus on	20	49	87
belong to	13	32	100
listen to	11	27	50
be about	9	22	1
learn from	9	22	4
speak with	9	22	5
switch to	8	20	57

Prepositional type 1	Raw	Norm	Faithfulness
identify with	5	11	22
speak to	5	11	9
come to	4	9	24
look into	3	7	10
think of	3	7	21
begin with	2	5	6
believe in	2	5	5
come into	2	5	12
differ from	2	5	11
make of	2	5	3
speak of	2	5	3
apply for	1	2	4
compare to	1	2	4
differ to	1	2	6
follow from	1	2	4
go into	1	2	3
go through	1	2	3
go for	1	2	3
reflect on	1	2	3
start with	1	2	11
Total	207	469	Mean= 28

Prepositional type 1	Raw	Norm	Faithfulness
talk about	8	20	12
reach to	7	17	19
speak in	7	17	4
distinguish between	6	15	12
talk with	6	15	9
come with	5	12	15
help in	5	12	17
apply on	4	10	11
ask about	4	10	10
apply to	3	7	8
apply for	2	5	6
begin by	2	5	17
know about	2	5	3
relate to	2	5	7
start with	2	5	13
think of	2	5	9
work with	2	5	12
apply in	1	2	3
apply with	1	2	3
ask for	1	2	3
be for	1	2	0
begin as	1	2	8
believe in	1	2	8
change from	1	2	4
differ in	1	2	3
do with	1	2	1
go through	1	2	3
learn with	1	2	0
speak about	1	2	1
start by	1	2	6
switch with	1	2	7
switch from	1	2	7
think about	1	2	4
work on	1	2	6
Total	324	790	Mean= 18

Phrasal prepositional type 1

Phrasal prepositional type 1	Raw	Norm	Faithfulness
find out about	2	5	2
look out for	1	2	3
Total	3	7	Mean= 3

Phrasal prepositional type 1	Raw	Norm	Faithfulness
come up with	2	5	6
find out about	1	2	4
Total	3	7	Mean= 5

Complex copular

Complex copular	Raw	Norm	Faithfulness
consider	24	54	48
make	24	54	38
see	24	54	32
use	19	43	7
define	17	39	26
know	13	29	33
find	11	25	12
view	9	20	50
describe	6	14	13
think	6	14	43
read	4	9	11
say	4	9	16
believe	4	9	10
claim	3	7	13
help	3	7	27
identify	2	5	9
reflect	2	5	6
show	2	5	4
teach	2	5	8
give	1	2	2
learn	1	2	2
note	1	2	4
understand	1	2	4
total	183	415	Mean= 18

Complex copular	Raw	Norm	Faithfulness
call	30	73	86
make	20	49	32
consider	17	41	100
help	9	22	30
know	9	22	15
use	6	15	1
say	4	10	3
chain	1	2	6
see	1	2	3
show	1	2	1
Total	98	239	Mean= 28

Complex transitive

Complex transitive	Raw	Norm	Faithfulness
reflect	6	14	18
find	1	2	1
Total	7	16	Mean=10

Complex transitive	Raw	Norm	Faithfulness
Total	0	0	Mean= 0

Ditransitive/ Double object

Ditransitive/ Double object	Raw	Norm	Faithfulness
allow	22	50	56
ask	9	20	56
give	14	32	31
lead	1	2	6
teach	4	9	16
Total	50	113	Mean= 33

Ditransitive/ Double object	Raw	Norm	Faithfulness
ask	23	56	61
give	15	37	31
tell	15	37	71
show	10	24	15
teach	2	5	6
Total	65	159	Mean= 37

Prepositional type 2a

Prepositional type 2a	Raw	Norm	Faithfulness
give to	5	11	11
allow for	3	7	8
Total	8	18	Mean= 10

Prepositional type 2a	Raw	Norm	Faithfulness
show to	3	7	4
show for	1	2	2
give to	9	22	18
Total	13	32	Mean= 8

Prepositional type 2b

Prepositional type 2b	Raw	Norm	Faithfulness
write for	1	2	2
write to	1	2	2
Total	2	5	Mean= 2

Prepositional type 2b	Raw	Norm	Faithfulness
seem to	6	15	18
say to	4	10	3
seem for	2	5	6
speak to	1	2	1
Total	13	32	Mean= 7

Prepositional type 3

Prepositional type 3	Raw	Norm	Faithfulness
take sth into account	6	14	13
take part in	4	9	9
take sth into consideration	3	7	7
take control of	1	2	2
Total	14	32	Mean= 8

Prepositional type 3	Raw	Norm	Faithfulness
take care of	6	15	25
switch attention to	2	5	14
make use of	1	2	2
give attention to	1	2	2
Total	10	24	Mean=11

Prepositional type 4a

Prepositional type 4a	Raw	Norm	Faithfulness
involve in	5	11	12
provide with	2	5	11
ask for	1	2	6
Total	8	18	Mean= 10

Prepositional type 4a	Raw	Norm	Faithfulness
call by	2	5	6
call with	1	2	3
want from	1	2	3
tell about	1	2	5
Total	5	12	Mean= 4

Prepositional type 4b

Prepositional type 4b	Raw	Norm	Faithfulness
base on	33	75	77
add to	22	50	59
apply to	11	25	46
compare to	11	25	46
combine with	8	18	33
provide for	5	11	28
compare with	5	11	21
attach to	4	9	24
involve in	3	7	7
provide to	2	5	11
change to	2	5	4
know for	2	5	5
apply in	2	5	8
make into	1	2	2
base around	1	2	2
say for	1	2	4

Prepositional type 4b	Raw	Norm	Faithfulness
relate to	21	51	75
based on	17	41	100
apply on	11	27	31
distinguish from	5	12	10
apply to	4	10	11
connect with	4	10	31
speak with	3	7	2
connect to	3	7	23
organize into	3	7	25
learn in	2	5	1
tell about	2	5	10
learn from	1	2	1
learn through	1	2	1
show through	1	2	1
distinguish into	1	2	2
give with	1	2	2

Prepositional type 4b	Raw	Norm	Faithfulness
relate to	1	2	5
Total	114	259	Mean= 22

Prepositional type 4b	Raw	Norm	Faithfulness
change with	1	2	4
take from	1	2	4
name after	1	2	4
identify with	1	2	5
Total	84	205	Mean= 17

Phrasal prepositional type 2

Phrasal prepositional type 2	Raw	Norm	Faithfulness
make up of	2	5	3
add on to	1	2	3
Total	3	7	Mean= 3

Phrasal prepositional type 2	Raw	Norm	Faithfulness
Total	0	0	Mean= 0

Phrasal prepositional type 3

Phrasal prepositional type 3	Raw	Norm	Faithfulness
lead on to	1	2	6
Total	1	2	Mean= 6

Phrasal prepositional type 3	Raw	Norm	Faithfulness
Total	0	0	Mean= 0

Appendix 8: Learners' errors and raters' judgement

	Verb		Rater 1	Rater 2
1	accord	1- Vocabulary differs *according* the person's choice.	Y	Y
2	activate	1- We *activate to* one, and we hold back of parsing.	Y	Y
3	affect	1- People *affected with* this type.	Y	Y
4	appear	1- The audiolingual method *appear by* the need to teach the soldiers.	Y	Y
		2- After the audiolingual method *appear another method.*	Y	Y
		3- To his principle *appear unfriendly* relationship.	Y	Y
5	ask	1- The teacher *asks* the learner will remember.	Y	Y
		2- Learning by *asking for and from* the teacher.	Y	Y
6	become	1- Interference *become under* the issue of transfer.	Y	Y
		2- This method *become as an opposite* to audiolingual method.	Y	Y
		3- The teacher's purpose has not *become to impose* limits and boundaries.	Y/N	Y
		4- It will *become relies* on himself.	Y	Y
7	come	1- His friends *come in* the home.	Y	Y
8	Denote	1- We cannot refer and *denote to* these creatures.	Y	Y
		2- Morning star have a different meaning but *denote to* the same object.	Y	Y
		3- It will not *denote to* object.	Y	Y
		4- The relevant words are by no means as its *denote.*	Y	Y
		5- Some words do not have a physical *denote* like love and hate.	Y	Y
		6- Noun have different meaning but the same *denote*.	Y	Y
		7- We have noun does not *denote to* physical object like love and hate.	Y	Y
9	depend	1- Our words choices *depend in* different things.	Y	Y
		2- The choice of words differs *depending to* when you are speaking.	Y	Y
10	Determine	1- Acquiring language is innate *determine.*	Y	Y
11	distinguish	1- We can *distinguish* of the real world.	Y	Y
		2- Vocabulary can *distinguish social class* into categories.	N	N

	Verb		Rater 1	Rater 2
		3- The vocabulary *distinguish* people to highest class and lowest class.	Y	Y
12	exist	1- Some information *is exists the knowledge* in human from birth.	Y	Y
13	focus	1- It *focus in* visuals and speaking tasks.	Y	Y
		2- Teachers *focus how* to teach.	Y	Y
14	get	1- Our intuition *get use* to possible meaning.	Y	Y
		2- Words in some case *get errors* while speech.	Y	Y
15	help	1- It helps the child to segmentation.	Y	Y
16	learn	1- She cannot *learning with* themselves.	Y	Y
		2- They should to *learn with* people.	Y	Y
		3- The interaction with others results *learning language.*	Y	Y
		4- Experience leads to *learning strengthening* particular connections.	Y	Y
		5- Imagery *learn the students* to use imagination.	Y	Y
		6- Student use imagination to *learn fluent.*	Y	Y
		7- We create new things to *learn with it*.	N	N
		8- Learn him how to read.	Y	Y
		9- Child *learns of this words* by the process of naming.	Y	Y
		10- The child learn many of word from process of naming.	Y	Y
17	listen	1- Sara prefer *listen the lecture* when she study.	Y	Y
18	make	1- Avoidance is want to *make avoid* the subject.	Y	Y
		2- If they feel a point is not important, they *make it avoid.*	Y	Y
		3- He goes back to the passage because he *make wrong.*	Y	Y
		4- They use “tend out” which *make low structure.*	Y	Y
		5- Come to clause boundaries and *make it off.*	Y	Y
		6- We *make automatically moving* to alternative interpretation.	Y	Y
		7- They *make experiment* about question.	Y	Y
		8- Make us more carefully.	Y	Y
		9- A test that was *made to subject* to the right word.	Y	Y
19	marry	1- The men must *marry in* outside tribe.	Y	Y

	Verb		Rater 1	Rater 2
20	occur	1- Does semantic processing *occurs the moment* of hearing the words.	Y	Y
21	organize	1- Approach in three properties which are *organize* the linguistics data in various classes.	Y	Y
22	practice	1- The mother trying or *practice* her child to speak.	Y	Y
23	produce	1- They can comprehend the speech but they *can't produce*.	Y/N	Y
24	pronounce	1- The upper class are more clear when they *pronounce*.	Y/N	Y
25	reach	1- When we *reach at* the end of the sentence.	Y	Y
26	refer	1- Interference *refers speakers* and writers applying knowledge.	Y	Y
		2- Types of learning is solving problems which *refers as* thinking.	Y	Y
		3- Some noun doesn't refer noun.	Y	Y
		4- This *refers that* at home use informal forms.	Y	Y
27	relate	1- Vernacular form *relate to pronunciation* and vocabulary.	Y	Y
		2- There are some nouns *do not relate to* the world.	Y	Y
		3- There are different words *that relate to the* same thing.	N	N
28	seem	1- This type of Dyslexia *seems that* the visual direct route is not damaged.	Y	Y
29	see	1- We can *see Ray's feeling* through his words.	N	N
		2- You will *see different* in vocabulary.	Y	Y
30	show	1- Differences in sociolinguistics *that are shown* a lot of things as social class.	Y	Y
		2- The complexity of language may *show from* a simple development process.	N	N
		3- Maybe also she want *to show respect to not speak* and use the language.	Y	Y
31	solve	1- Multiple learning, chain learning, to final *solving* problem.	Y	Y
32	speak	1- They don't speak each other.	Y	Y
		2- We don't *speak our bosses* in the same way as we speak.	Y	Y
		3- Men do not *speak in exactly way* as speech other in day community.	Y	Y
		4- With your friend you will *speak free*.	Y	Y
		5- Women and men do not *speak different* as each other.	Y	Y
		6- The upper class *speak different* than the lower class.	Y	Y

	Verb		Rater 1	Rater 2
		7- People of different social class do not *speak the same way.*	N	N
		8- It is not the same way as you speaking at your mother at home.	Y	Y
33	start	1- They start produce words.	Y	Y
34	talk	1- A bank manager doesn't *talk the same way* the office cleaner talk.	Y	Y
		2- Tayan still talk with his first language.	Y	Y
		3- Each of them talk with his own language.	Y	Y
35	think	1- Which thought to contain all and only principles.	Y	Y
		2- At which point they think the clock had occurred.	Y	Y
		3- The mother is used here think about her prestigious.	Y	Y
36	use	1- The words *used between* men are longer.	N	N
37	want	1- Compensating is *want to extra* thing such as when you ask the teacher.	Y	Y